

UNIVERSIDADE FEDERAL DO PARÁ
INSTITUTO DE TECNOLOGIA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA

**ANÁLISE DOS IMPACTOS HARMÔNICOS NA QUALIDADE DA ENERGIA
ELÉTRICA UTILIZANDO KDD – ESTUDO DE CASO NA UNIVERSIDADE
FEDERAL DO PARÁ**

WATERLOO FERREIRA DA SILVA

UFPA / ITEC / PPGEE
Campus Universitário do Guamá
Belém-Pará-Brasil

2019

UNIVERSIDADE FEDERAL DO PARÁ
INSTITUTO DE TECNOLOGIA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA

WATERLOO FERREIRA DA SILVA

**ANÁLISE DOS IMPACTOS HARMÔNICOS NA QUALIDADE DA ENERGIA
ELÉTRICA UTILIZANDO KDD – ESTUDO DE CASO NA UNIVERSIDADE
FEDERAL DO PARÁ**

DM 600/2019

UFPA / ITEC / PPGEE
Campus Universitário do Guamá
Belém-Pará-Brasil

2019

UNIVERSIDADE FEDERAL DO PARÁ
INSTITUTO DE TECNOLOGIA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA

WATERLOO FERREIRA DA SILVA

**ANÁLISE DOS IMPACTOS HARMÔNICOS NA QUALIDADE DA ENERGIA
ELÉTRICA UTILIZANDO KDD – ESTUDO DE CASO NA UNIVERSIDADE
FEDERAL DO PARÁ**

Dissertação de Mestrado
apresentada à coordenação do Programa
de Pós-Graduação em Engenharia
Elétrica (PPGEE) do Instituto de
Tecnologia da Universidade Federal do
Pará (UFPA) como parte dos requisitos
finais para obtenção do grau de Mestre
em Engenharia Elétrica.

UFPA / ITEC / PPGEE
Campus Universitário do Guamá
Belém-Pará-Brasil

2019

**Dados Internacionais de Catalogação na Publicação (CIP) de acordo com ISBD
Sistema de Bibliotecas da Universidade Federal do Pará
Gerada automaticamente pelo módulo Ficat, mediante os dados fornecidos pelo(a) autor(a)**

F383a Ferreira da Silva, Waterloo
Análise dos impactos harmônicos na qualidade da energia elétrica utilizando KDD – estudo de caso na Universidade Federal do Pará / Waterloo Ferreira da Silva. — 2019.
120 f. : il.

Orientador(a): Prof^a. Dra. Maria Emília de Lima Tostes
Dissertação (Mestrado) - Programa de Pós-Graduação em Engenharia Elétrica, Instituto de Tecnologia, Universidade Federal do Pará, Belém, 2019.

1. Sistemas de Energia Elétrica - Controle de Qualidade. 2. Mineração de Dados. 3. Analise Harmônicos . 4. Estudo de Caso Universidade Federal do Pará. 5. Naive Bayes. I. Título.

CDD 621.310285

**“ANÁLISE DOS IMPACTOS HARMÔNICOS NA QUALIDADE DA
ENERGIA ELÉTRICA UTILIZANDO KDD-ESTUDO DE CASO NA
UNIVERSIDADE FEDERAL DO PARÁ”**

AUTOR: **WATERLOO FERREIRA DA SILVA**

DISSERTAÇÃO DE MESTRADO SUBMETIDA À BANCA EXAMINADORA APROVADA PELO
COLEGIADO DO PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA, SENDO
JULGADA ADEQUADA PARA A OBTENÇÃO DO GRAU DE MESTRE EM ENGENHARIA
ELÉTRICA NA ÁREA DE SISTEMAS DE ENERGIA ELÉTRICA.

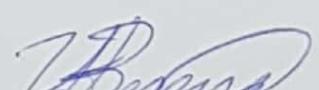
APROVADA EM: 18/03/2019

BANCA EXAMINADORA:



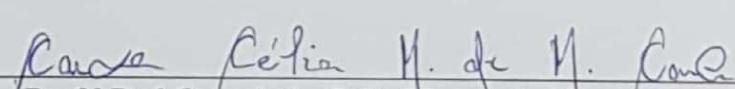
Prof.ª Dr.ª Maria Emilia de Lima Tostes

(Orientadora – PPGEE/UFPA)



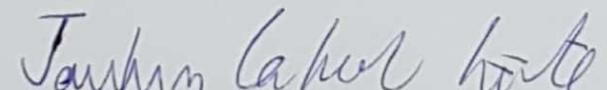
Prof. Dr. Ubiratan Holanda Bezerra

(Avaliador Interno – PPGEE/UFPA)



Prof.ª Dr.ª Carminda Celia Moura de Moura Carvalho

(Avaliadora Externa ao Programa – FEEB/UFPA)



Prof. Dr. Jandecy Cabral Leite

(Avaliador Externo – ITEGAM)

VISTO:

Prof.ª Dr.ª Maria Emilia de Lima Tostes

(Coordenadora do PPGEE/ITEC/UFPA)

AGRADECIMENTOS

A Deus por me guiar pelos caminhos da vida.

A Universidade Federal do Pará - UFPA.

Ao Instituto de Tecnologia e Educação Galileo da Amazônia – ITEGAM.

Aos Professores, Dr. Ádamo Lima de Santana, Dra. Maria Emília de Lima Tostes e Dr. Jandecy Cabral Leite, pela condução e orientação com excelência durante o curso.

Ao amigo M.Sc. Rildo de Mendonça Nogueira, pelo apoio e ajuda incomensurável durante a realização deste trabalho.

A Universidade Paulista UNIP (Campus de Manaus), colega de mestrado pelo incentivo nesses meses de estudo e na preparação desta dissertação.

Aos professores do curso.

Aos meus pais Sergio e Rosária, que me incentivam em todos os momentos da vida.

Aos meus filhos pelo apoio, compreensão e dedicação.

Aos meus queridos irmãos pelo apoio, compreensão e união.

A todos que, direta e indiretamente, contribuíram para a realização deste trabalho.

Dedicatória

Dedico este trabalho em primeiro lugar a Deus por me guiar durante toda a caminhada da vida, aos meus pais Sergio Ferreira da Silva e Rosária Maria da Silva que em todos os momentos da vida me incentivaram a ir além, aos meus irmãos Sergio e Adriana pelo seu exemplo de vida, as pessoas muito especiais em minha vida, meus filhos Thiago, Waterloo Filho, Lucas, Vinicius, Gabriel, João Paulo.

‘1. Provérbios de Salomão, filho de Davi, rei de Israel:

2. Para se conhecer a sabedoria e a instrução; para se entenderem as palavras de inteligência;

3. para se instruir em sábio procedimento, em retidão, justiça e equidade;

4. para se dar aos simples prudência, e aos jovens conhecimento e bom siso.

5. Ouça também, o sábio e cresça em ciência, e o entendido adquira habilidade,

6. para entender provérbios e parábolas, as palavras dos sábios, e seus enigmas.

7. O temor do Senhor é o princípio do conhecimento; mas os insensatos desprezam a sabedoria e a instrução.

8. Filho meu, ouve a instrução de teu pai, e não deixes o ensino de tua mãe.

9. Porque eles serão uma grinalda de graça para a tua cabeça, e colares para o teu pescoço.

10. Filho meu, se os pecadores te quiserem seduzir, não consintas.’

(Provérbios 1)

LISTA DE FIGURAS

Figura 1 – Onda deformada e suas componentes harmônicas.....	36
Figura 2 - Harmônico de sequência zero.....	38
Figura 3 - Processo de KDD.....	52
Figura 4 - Processo de seleção de dados.....	53
Figura 5 - Processo de Armazenamento em Banco de Dados e Sistema Operacional.....	55
Figura 6 - Classificação como a tarefa de mapear um conjunto de atributos x no seu rótulo de classe y	56
Figura 7 - Processo de mineração de dados.....	58
Figura 8 - Fase de Interpretação do KDD.....	59
Figura 9 - Esquema de um classificador.....	60
Figura 10 - Partição do espaço amostral Ω	62
Figura 11 - Modelo de um Naive Bayes.....	67
Figura 12 - Imagem ilustrativa do Analisador de QEE HIOKI PW3198.....	72
Figura 13 - Seleção dos Dados do Analisador para Exportação.....	73
Figura 14 - Dados discretizados harmônica 7 básico 1.....	80
Figura 15 - Gráfico ROC da fase A.....	84
Figura 16 - Gráfico ROC da fase B.....	85
Figura 17 - Gráfico ROC da fase C.....	86
Figura 18 - Naive Bayes fase A.....	88
Figura 19 - Densidade de probabilidade dos intervalos discretizados do THDv em relação a corrente do 3º harmônico.....	89
Figura 20 - Densidade de probabilidade dos intervalos discretizados do THDv em relação a corrente do 7º harmônico.....	90
Figura 21 - Densidade de probabilidade dos intervalos discretizados do THDv em relação a corrente do 5º harmônico.....	91
Figura 22 - Densidade de probabilidade dos intervalos discretizados do THDv em relação a corrente do 9º harmônico.....	92
Figura 23 - Densidade de probabilidade dos intervalos discretizados do THDv em relação a corrente do 5º harmônico.....	93
Figura 24 - Densidade de probabilidade dos intervalos discretizados do THDv em relação a corrente do 7º harmônico.....	94

Figura 25 - Densidade de probabilidade dos intervalos discretizados do THDv em relação a corrente do 5º harmônico.	95
Figura 26 - Densidade de probabilidade dos intervalos discretizados do THDv em relação a corrente do 7º harmônico.	96
Figura 27 - Naive Bayes fase B.	97
Figura 28 - Densidade de probabilidade dos intervalos discretizados do THDv em relação a corrente do 3º harmônico.	98
Figura 29 - Densidade de probabilidade dos intervalos discretizados do THDv em relação a corrente do 7º harmônico.	99
Figura 30 - Densidade de probabilidade dos intervalos discretizados do THDv em relação a corrente do 3º harmônico.	100
Figura 31 - Densidade de probabilidade dos intervalos discretizados do THDv em relação a corrente do 9º harmônico.	101
Figura 32 - Densidade de probabilidade dos intervalos discretizados do THDv em relação a corrente do 11º harmônico.	102
Figura 33 - Densidade de probabilidade dos intervalos discretizados do THDv em relação a corrente do 7º harmônico.	103
Figura 34 - Densidade de probabilidade dos intervalos discretizados do THDv em relação a corrente do 5º harmônico.	104
Figura 35 - Densidade de probabilidade dos intervalos discretizados do THDv em relação a corrente do 7º harmônico.	105
Figura 36 - Naive Bayes fase C.	106
Figura 37 - Densidade de probabilidade dos intervalos discretizados do THDv em relação a corrente do 7º harmônico.	107
Figura 38 - Densidade de probabilidade dos intervalos discretizados do THDv em relação a corrente do 11º harmônico.	108
Figura 39 - Densidade de probabilidade dos intervalos discretizados do THDv em relação a corrente do 3º harmônico.	109
Figura 40 - Densidade de probabilidade dos intervalos discretizados do THDv em relação a corrente do 7º harmônico.	110
Figura 41 - Densidade de probabilidade dos intervalos discretizados do THDv em relação a corrente do 5º harmônico.	111
Figura 42 - Densidade de probabilidade dos intervalos discretizados do THDv em relação a corrente do 7º harmônico.	112

LISTA DE ABREVIATURAS E SIGLAS

ANEEL	Agencia Nacional de Energia Elétrica
KDD	Knowledge Discovery in Databases
PRODIST	Procedimentos de Distribuição de Energia Elétrica no Sistema Elétrico Nacional
DM	Data mining - Mineração de Dados
IC	Inteligência Computacional
SEP	Sistema Elétrico de Potencia
PAC	Ponto de Acoplamento Comum
IEC	Comissão Eletrotécnica
THD	Distorção Total de Harmônicas
QEE	Qualidade da Energia Elétrica
ONS	Operador Nacional do Sistema Elétrico
DIT	Demais Instalações de Transmissão
FP	Fator de Potência
IEC	International Electro technical Commission
AD	Árvore de Decisão
RNAs	Redes Neurais Artificiais
ONS	Operador Nacional do Sistema Elétrico
CSV	<i>Comma separated values</i>
SGBD	Sistemas de Gerenciamento de Bancos de Dados
SOM	Sistemas Operacionais de Máquinas
THDu	Distorção Harmônica Total de Tensão
THDv	Distorção Harmônica Total de Tensão
TDHi	Distorção Harmônica Total de Corrente
WEKA	Waikato Environment for Knowledge Analysis
SQL	Structured Query Language
FP	Falso Positivo
TP	Verdadeiro Positivo

LISTA DE TABELAS

Tabela 1 - Ordem, frequência e sequência das harmônicas.	37
Tabela 2 - Níveis de compatibilidade para as tensões harmônicas individuais em redes de baixa (BT) e média tensão (MT).....	42
Tabela 3 - Níveis de planejamento para tensões harmônicas individuais em redes de média .	43
Tabela 4 - Limites de distorção de tensão.	44
Tabela 5 - Limites de distorção de corrente para sistemas de 120 V a 69 KV.....	45
Tabela 6 - Limites de distorção de corrente para sistemas com classificação acima de 69 kV até	45
Tabela 7 - Limites de distorção de corrente para sistemas classificados > 161 Kv.....	46
Tabela 8 - Terminologia	47
Tabela 9 - Expressões para o cálculo das grandezas DITh%, DTT%, DTTp%, DTTi% e DTT3% são.	48
Tabela 10 - Os limites para as distorções harmônicas totais.	48
Tabela 11 - Os Alimentadores da Universidade Federal do Pará.....	70
Tabela 12 - Características dos analisadores de QEE.....	71
Tabela 13 - Especificação técnica dos dados dos arquivos CSV.	74
Tabela 14 - Estrutura dos campos das tabelas no SGDB.	75
Tabela 15 - Transformação dos dados.	76
Tabela 16 - Nomenclaturas dos atributos utilizados nas reduções da fase A.....	77
Tabela 17 - Nomenclaturas dos atributos utilizados nas reduções da fase B.	78
Tabela 18 - Nomenclaturas dos atributos utilizados nas reduções da fase C.	79
Tabela 19 - Matriz de Confusão para duas Classes.	81
Tabela 20 - Termos utilizados no desempenho do classificador.	82
Tabela 21 - Matriz confusão da fase A.	83
Tabela 22 - Acurácia da Fase A.....	84
Tabela 23 - Matriz confusão da fase B.....	84
Tabela 24 - Acurácia da Fase B.....	85
Tabela 25 - Matriz confusão da fase C.....	85
Tabela 26 - Acurácia da Fase C.....	86
Tabela 27 - Sumário de análises de fases - Mineração de dados.....	87

LISTA DE EQUAÇÕES

Equação 1	Cálculo da Distorção Harmônica Total de Tensão.....	40
Equação 2	Cálculo da Distorção Harmônica Individual de Tensão de Ordem h.....	40
Equação 3	Fator K_{hvs}	41
Equação 4	Cálculo soma das probabilidades.....	61
Equação 5	Axiomas (1) e (2) combinados.....	61
Equação 6	Axiomas (1) e (3) combinados.....	61
Equação 7	Cálculo da Classe de maior probabilidade para classificador Naive Bayes.....	63
Equação 8	Cálculo de probabilidade de diferentes hipóteses.....	64
Equação 9	Probabilidade de observemos simultaneamente eventos A e B.....	64
Equação 10	Combinando equações (5) e (6).....	64
Equação 11	Teorema de Bayes	65
Equação 12	Cálculo de probabilidade da inferência aplicada na rede	66
Equação 13	Calculo da Acurácia	81
Equação 14	Calculo da Sensitividade (Recall).....	81
Equação 15	Calculo da Precisão	81
Equação 16	Cálculo da Especificidade	82

RESUMO

O presente trabalho apresenta uma análise de dados relacionados a Qualidade da Energia Elétrica (QEE), a crescente utilização de cargas não lineares, equipamentos baseados na eletrônica de potência nas instalações residenciais, comerciais e industriais estão contribuindo para o aumento significativo dos níveis de distorção harmônica de corrente e, conseqüentemente de tensão, conforme observado no sistema brasileiro de distribuição de energia elétrica.

Foi desenvolvido no Brasil, os procedimentos de distribuição no sistema elétrico nacional (PRODIST), criado e desenvolvido pela Agencia Nacional de Energia Elétrica (ANEEL). O PRODIST tem o objetivo de normatizar e padronizar as atividades relacionadas à distribuição de energia, incluindo os padrões de qualidade do produto.

Com objetivo de monitorar a qualidade do produto “energia elétrica” através do conteúdo harmônico gerado pela rede elétrica da instituição, propõe-se uma metodologia para as analise utilizando técnicas de inteligência computacional (IC) e mineração de dados para análise dos dados coletados por medidores de qualidade de energia instalados nos principais setores desta instituição e no ponto de acoplamento comum do consumidor e conseqüentemente estabelecer a relação entre as correntes harmônicas das cargas não lineares com a distorção harmônica no ponto de acoplamento comum.

Aplicou-se o processo do KDD, incluindo coleta, seleção, limpeza, integração, transformação e redução, mineração, interpretação e avaliação dos dados, com objetivo de monitorar a qualidade do produto “energia elétrica” através do conteúdo harmônico gerado pela rede elétrica da instituição de ensino. Na fase de mineração de dados “Data Mining”, utilizou-se o classificador Naive Bayes.

Os resultados obtidos evidenciaram que o processo KDD possui aplicabilidade na análise da Distorção Harmônica Total de Tensão no Ponto de Acoplamento Comum podendo ser aplicado em qualquer área comercial, residencial e industrial.

Palavras-chave: KDD, Naive Bayes, Harmônicos, de mineração de dados, sistemas de potência.

ABSTRACT

The present work presents an analysis of data related to Power Quality (PQ), the increasing use of nonlinear loads, equipment based on power electronics in residential, commercial and industrial installations are contributing to the significant increase in the levels of harmonic distortion of current and, consequently, of voltage, as observed in the Brazilian electricity distribution system.

It was developed in Brazil, the distribution procedures in the national electricity system (PRODIST), created and developed by the National Electric Energy Agency (ANEEL). PRODIST aims to standardize and standardize activities related to energy distribution, including product quality standards.

In order to monitor the quality of the product "electric energy" through the harmonic content generated by the electric network of the institution, a methodology is proposed for the analysis using computational intelligence (CI) and data mining techniques to analyze the data collected by meters of energy quality installed in the main sectors of this institution and at the point of common coupling of the consumer and consequently establish the relationship between the harmonic currents of the nonlinear loads with the harmonic distortion at the common coupling point.

The KDD process was applied, including the collection, selection, cleaning, integration, transformation and reduction, mining, interpretation and evaluation of the data, in order to monitor the quality of the product "electric energy" through the harmonic content generated by the electric grid. educational institution. In the "Data Mining" data mining phase, the Naive Bayes classifier was used.

The obtained results showed that the KDD process has applicability in the analysis of the Total Harmonic Distortion of Voltage at the Common Coupling Point and can be applied in any commercial, residential and industrial area.

Keywords: KDD, Naive Bayes, Harmonic, Data Mining, Power Quality.

CAPÍTULO I

INTRODUÇÃO

1.1 Contextualização

O tema estudado neste presente trabalho está relacionado com a qualidade de energia elétrica (QEE). Com o avanço na eletrônica de potência ocorreu um aumento significativo da quantidade de dispositivos eletrônicos, que constituem carga não-lineares, estão presentes nos sistemas elétricos em universidades, indústrias, áreas comerciais e residenciais, são bastante variados tais como computadores, iluminação de descarga gasosa, conversores de frequência, forno de micro ondas, máquina de lavar louças e televisores “led, plamas”, estes equipamentos oferecem uma série de benefícios para os consumidores e para o setor elétrico, melhorando a produtividade e a eficiência de processos ou sistemas, no entanto, em contrapartida, a proliferação de tais dispositivos provoca o surgimento de harmônicos na rede elétrica, ocasionando por sua vez, inúmeros efeitos indesejados, como redução da vida útil de equipamentos, interrupção de processos industriais, entre outros.

Este tema vem sendo, nos últimos anos objeto de preocupação e de esforços coordenados de diagnósticos e de análise para que se atinja um maior entendimento possível sobre o mesmo. Protagonistas destes estudos têm sido tanto concessionárias de energia elétrica, quanto grandes consumidores, bem como universidades, centros de pesquisa e fabricantes de equipamentos.

Em virtude disso a matéria tem merecido ampla divulgação através de um número inevitável de artigos, informes técnicos e comentários em publicações nacionais e internacionais. O estudo, a caracterização, a compreensão, em resumo, a abordagem de alguns dos fenômenos que afetam a QEE, já vem apresentando avanços significativos.

Os harmônicos injetados no sistema elétrico pelas cargas não lineares produzem efeitos: nos próprios sistemas elétricos de potência e nas cargas elétricas conectadas a estes, assim como nos sistemas de comunicações.

Os efeitos dos harmônicos em sistemas de potência são prejudiciais e entre eles pode-se destacar os seguintes:

- a) Harmônicas de corrente aumentam as perdas no cobre, principalmente devido ao efeito pelicular que implica na redução efetiva da área condutora à medida que se eleva a frequência da componente de corrente;
- b) Harmônicas de tensão aumentam as perdas no ferro;
- c) Normalmente, devido amplitudes reduzidas das harmônicas, este aumento de perdas não é excessivo;
- d) Ressonâncias podem ser excitadas pelas componentes harmônicas, resultando em níveis excessivos de tensão e/ou corrente;
- e) Reatância capacitiva se reduz com a frequência, aumentando-se as correntes relativas às harmônicas presentes na tensão;
- f) Harmônicos elevam valores eficazes da corrente que resultam em aumento das perdas nos condutores;
- g) Efeito pelicular que restringe a área útil do condutor para componentes de elevadas frequências;
- h) Equipamentos eletrônicos - podem ser muito sensíveis à distorção harmônica da tensão, principalmente quando a utilizam como referência para sua operação, e/ou, execução de determinada tarefa;
- i) Fusíveis - devido harmônicas e inerente aumento da corrente eficaz, resultando em maior aquecimento, poderá levar à redução da vida útil dos mesmos, ou até mesmo, à uma operação inadequada.

Em face do exposto, é de suma importância a priorização do tema, sobretudo em projetos de novas unidades industriais e comerciais, bem como nos de modernização das já existentes. Não se deve, contudo, deixar de observar que, em geral, soluções em qualidade de energia só são passíveis de serem alcançados, quando consumidores, fornecedores de energia e fabricantes de equipamentos conscientizam-se da necessidade de entendimento mútuo para tal.

O presente trabalho de pesquisa apresenta um estudo para diagnosticar a qualidade de energia elétrica afetada pelos distúrbios provenientes das distorções harmônicas analisando o indicador total de distorção harmônica (THD) no ponto de acoplamento comum - PAC entre uma universidade e concessionária, através da utilização técnicas de mineração de dados mais especificamente com o classificador Naive Bayes.

1.2 Justificativa da Pesquisa

Um dos principais distúrbios que afetam a qualidade da energia elétrica são as distorções harmônicas, as quais são geradas por cargas não-lineares conectadas a rede elétrica estas, provocam a circulação das correntes harmônicas que geram tensões harmônicas através da impedância da rede, tendo como resultado a deformação da tensão de alimentação. Uma carga produtora de harmônicos pode afetar outras cargas se uma distorção significativa é produzida.

A identificação e determinação servem para avaliar, eliminar, mitigar ou amenizam a degradação na qualidade da energia elétrica.

No Brasil, a Agência Nacional de Energia Elétrica (ANEEL) desenvolveu os procedimentos de distribuição de energia (ANEEL-PRODIST, 2018), que é um conjunto de documentos composto de oito módulos voltados para a normatização e padronização das atividades técnicas relacionadas ao funcionamento e desempenho dos sistemas de distribuição de energia elétrica, dentre os quais no 8º módulo são estabelecidos limites de distorção harmônica total e individual de tensão para um sistema de distribuição qualquer.

A presença de harmônicos gera impactos econômicos de grandes proporções tanto para os consumidores de campus universitários e residenciais quanto para a cadeia produtiva (comércio e indústria). Fica assim evidente a necessidade de métodos ou planos para eliminar ou atenuar o impacto das distorções harmônicas no sistema elétrico.

O referido estudo surge como uma alternativa para apoiar as políticas de prevenção através da avaliação e diagnóstico da qualidade de energia elétrica sob a ótica das distorções harmônicas, monitorando os indicadores da taxa de distorção harmônica por meio de um sistema de apoio a decisão utilizando técnicas estatísticas e mineração de dados mais especificamente o classificador de Naive Bayes.

1.3 OBJETIVOS

1.3.1 Objetivo Geral

Avaliar o conteúdo harmônico existente na rede elétrica de um Campus Universitário, através de parâmetros obtidos com medições feitas por equipamentos devidamente calibrados, utilizando técnicas estatísticas e mineração de dados.

1.3.2 Objetivos Específicos

- Apresentar um estudo sobre a qualidade de energia elétrica, tendo como base o indicador, a taxa de distorção harmônica (THD);
- Analisar as áreas envolvidas do campus, viabilizar uma avaliação do conteúdo harmônico presente em cada área, analisar os maiores impactos de harmônicos;
- Fornecer informações com base nos dados analisados que auxiliem na tomada de decisão e controle dos harmônicos gerados em cada área do Campus Universitário, evitando penalidades devido a geração de harmônicos fora dos limites estabelecidos nas normas vigentes das agências reguladora;

1.4 Contribuição e Relevância do Estudo

Este trabalho visa auxiliar na tomada de decisão para resolução de alguns problemas relacionados a qualidade de energia elétrica (QEE) na rede elétrica de um Campus Universitário, onde o sistema possui perturbações harmônicas. O aumento dos níveis de distorção harmônica da tensão e da corrente são bastante danosos aos diversos equipamentos eletrônicos, a probabilidade de ocorrência de falhas nesses equipamentos, e, conseqüentemente, a interrupção do fornecimento de energia elétrica, queima de equipamentos, perdas econômicas, entre outros. Neste sentido, podem-se destacar ainda os seguintes aspectos e contribuições:

- Com a existência de distorções harmônicas na rede elétrica, sua avaliação tem sido um desafio perante os eventos que ocorrem na rede elétrica tanto, residenciais como comerciais e industriais;
 - Utilizar técnicas estatísticas de Mineração de Dados no apoio a análise da qualidade de energia elétrica em um Campus Universitário;
 - Analisar o perfil harmônico individual envolvido no Campus Universitário, conseqüentemente as contribuições no THD e PAC da referida Universidade;
 - Analisou-se o impacto harmônico de tensão a partir das correntes medidas provenientes das cargas não lineares que compõem um determinado sistema.

1.5 Delimitação da Pesquisa

Este estudo delimita-se a investigar as condições da QEE em um Campus Universitário, na avaliação para diagnosticar os impactos de harmônicos na rede elétrica de

tensão utilizando de inteligência computacional e mineração de dados, através da análise do THD, com base nas normas presentes no módulo 8 Revisão 8 (PRODIST/ANEEL, 2018), a IEEE Std. 519 – 2014 (AL-GARADI, KHAN *et al.*, 2016) e IEC 61000-3-6 (IEC/TR, 2008-02).

Realizou-se o monitoramento dos níveis de harmônicos em uma campanha de medição de uma semana, nos setores: Básico 1 e 2, Profissional, Saúde e na entrada da Universidade Federal do Pará (UFPA).

1.6 Estrutura da Dissertação

O presente trabalho de dissertação é composto por seis capítulos, desta forma estruturado:

O Capítulo 1 – Apresenta o tema através de uma Introdução, os objetivos gerais, e a estrutura da dissertação.

O Capítulo 2 – Apresenta o estado da arte com relação ao tema da dissertação.

O Capítulo 3 – Apresenta os aspectos que envolvem a Qualidade da Energia Elétrica, tais como: os principais fundamentos a respeito do fenômeno da distorção harmônica, conceitos, suas principais fontes geradoras, os efeitos causados nas redes elétricas e as principais normas associadas à mesma.

Capítulo 4 – Apresenta o processo de KDD, as fases do processo de descoberta de conhecimento em bases de dados.

Capítulo 5 – Apresenta o classificador Naive Bayes sua estrutura e aplicação, o qual foi utilizado na para gerar o estudo de caso desta dissertação.

Capítulo 6 – Apresenta o estudo de caso com aplicação da mineração de dados utilizando a técnica de Naive Bayes, a qual gera informações que dão suporte a tomada decisão.

Capítulo 7 - Apresenta as conclusões desta dissertação, assim como sugestões de como este trabalho pode ser futuramente melhorado e ampliado.

CAPÍTULO II

REVISÃO BIBLIOGRÁFICA

2.1 Estado da arte

A utilização de técnicas de inteligência computacional e mineração de dados estão sendo amplamente utilizadas nas mais diversas áreas do conhecimento e da pesquisa científica. Entre as várias aplicações, neste capítulo serão exibidas as aplicações nas áreas que possuem correspondência entre energia elétrica e inteligência computacional com uso de algoritmos de mineração de dados.

2.2 Trabalhos correlatos

No trabalho apresentado por (ZHANG e LU, 2002), os Classificadores Bayesianos são precisos nas tarefas de apreender utilizando bancos de dados. O tema desta pesquisa é muito utilizado nas prospecções de dados e aprendizagem automática. No entanto, na prática, as bases de dados são raramente completas. Isso afeta suas aplicações de mineração de dados do mundo real. Neste trabalho investiga-se os métodos de aprendizagem de quatro tipos bem conhecidos de classificadores Rede Bayesiana (RB) em bancos de dados incompletos.

Esses quatro classificadores são dos tipos RB: Naive Bayes, árvore aumentada Naive Bayes, RB aumentada Naive Bayes e RB geral, onde os dois últimos são aprendidos usando algoritmos baseados em análise de dependência que funcionam apenas na suposição de completude de banco de dados. Para permitir a manipulação de algoritmos desse tipo com dados perdidos, o artigo apresenta um método determinístico inovador para estimar a informação mútua (condicional) de bancos de dados incompletos, que pode ser usado para fazer testes de Independência Condicional (IC), uma etapa fundamental nos algoritmos baseados em análise de dependência. Os resultados experimentais mostram que o algoritmo é eficiente e confiável.

O trabalho desenvolvido por (TORRES, RUEDA *et al.*, 2006), apresenta um modelo para avaliar a QEE através do fenômeno denominado de afundamentos de tensão por meio de redes bayesianas. Este modelo considera os principais eventos que podem ser gerados pela presença de afundamentos de tensão nos sistemas industriais de energia, especialmente em equipamentos eletrônicos e sensíveis. As relações de causalidade estabelecidas entre as

principais causas e os efeitos finais consideram tanto o interno e as condições elétricas externas do sistema de energia industrial, tais como os efeitos ambientais.

No trabalho desenvolvido por (CESAR, VALDOMIRO *et al.*, 2006), foram utilizados alguns padrões baseados na transformada wavelet discreta com objetivo de detectar e identificar os distúrbios de baixa frequência, como flicker e harmônicos e distúrbios de alta frequência, como transitórios e afundamentos. A função wavelet é utilizada como base da função na detecção e identificação por causa de suas propriedades de localização, e de tempo de resposta de frequência e de informação. Com base nesses padrões, distúrbios de qualidade de energia são automaticamente classificados usando várias técnicas artificiais inteligentes: rede propagação neural (perceptron multicamadas), rede neural de Kohonen (auto mapa organizadora), Bayesiana (método estatístico linear) e Máquinas de Suporte Vetor (SVM).

No trabalho desenvolvido por (ASHEIBI, STIRLING *et al.*, 2006), o rápido aumento da tecnologia de computador e a disponibilidade de dados de monitoramento de qualidade de energia em larga escala agora devem motivar os provedores de serviços de rede de distribuição a tentar extrair informações que, de outra forma, permaneçam escondidas dentro dos dados gravados. Tais informações podem ser críticas para identificação e diagnóstico de problemas de perturbação de qualidade de energia, previsão de anormalidades ou falhas do sistema e alarmante de situações críticas do sistema. As ferramentas de mineração de dados são um candidato óbvio para auxiliar na análise de dados de monitoramento de qualidade de energia em larga escala.

No trabalho apresentado por (RAMOS, TORRES *et al.*, 2007), a aplicação de eletrônica de potência em sistemas industriais tem aumentado a utilização de uma ampla variedade de software de análise de confiabilidade para sistemas elétricos comerciais e industriais. No entanto, há eventos de qualidade de energia e de segurança que não são considerados por essas ferramentas. Este artigo descreve uma metodologia para a avaliação da confiabilidade de sistemas elétricos industriais usando redes Bayesianas, que incorpora as características de qualidade de energia e de segurança sob tais eventos nos curtos-circuitos, interrupções aleatórias e transientes elétricos. Esta metodologia é aplicada para a rede padrão proposto pelo IEEE 493 Livro de Ouro, para comparar os índices de confiabilidade do sistema com os obtidos por outros métodos. Este artigo, oferece uma ferramenta sólida e prática para a concepção de sistemas elétricos industriais.

No desenvolvimento de (CHENINE, NORDSTROM *et al.*, 2007), é apresentada uma pesquisa sobre o assunto de modelagem, desempenho, avaliação e preocupações relacionadas

aos sistemas de qualidade na concepção de tecnologias de informação e comunicação das plataformas utilizadas para a operação do sistema de potência e controle. O foco do projeto é sobre a forma como o desempenho e outros aspectos de qualidade de tais redes interagem com as tecnologias de sistemas de controle de energia modernas, como por exemplo sistemas de monitoramento e controle de área ampla (wide area monitoring and control - WAMC) o trabalho apresenta uma abordagem de avaliação e modelagem baseada em um conjunto de fatores teoricamente importantes e fundamentadas para avaliar o desempenho dos sistemas de controle de área ampla. Estes fatores de avaliação são utilizados para definir a semântica de modelagem necessárias para modelos aprimorados das plataformas de informação e tecnologia de comunicação. Os fatores são bem representados em diagramas de influência, permitindo uma análise da sua interdependência com, por exemplo, análise Bayesiana. Esta abordagem tem o potencial de oferecer suporte à tomada de decisão para a concepção de plataformas de informação e tecnologia de comunicação de alto desempenho, ao mesmo tempo, considerando outras qualidades do sistema.

No trabalho de (D'ANTONA, MUSCAS *et al.*, 2008) é proposto um novo procedimento, com base em uma abordagem bayesiana, destinada a fornecer informações aos operadores de rede de distribuição sobre a possível presença de cargas que produzem harmônicas, juntamente com uma indicação sobre a confiabilidade de tais informações. Dado que muitos instrumentos de medição são necessários para uma observação completa das quantidades elétricas, aqui se propõe uma técnica de estimação baseada no modelo de estado. Ela explora as poucas medições em tempo real e todas as outras informações disponíveis para estimar as correntes harmônicas injetadas pelas cargas não lineares. A robustez do processo no que diz respeito às possíveis variações na potência exigida pelas cargas e no seu comportamento da harmônica é investigada por meio de simulações de Monte Carlo.

Em (D'ANTONA, MUSCAS *et al.*, 2009), mostra que, ao longo dos últimos anos, com a liberalização do mercado da energia, a localização das fontes de distúrbios de qualidade de energia tornou-se um ponto nodal para os operadores e clientes. No mercado liberalizado, o conhecimento do status do sistema, juntamente com a incerteza que afeta os dados estimados, é necessário para operar as questões de gestão e de controle, tais como envio de energia e de coordenação da proteção. Nos sistemas de distribuição, seriam necessários muitos instrumentos de medição para obter uma observação completa das quantidades elétricos. Assim, as técnicas de estimativa de estado baseado em modelo pode ser implementado para explorar as poucas medições em tempo real a partir do campo e todas as outras informações disponíveis para

estimar as quantidades desejadas. No trabalho, o autor propõe uma nova abordagem para a estimativa das fontes harmônicas por meio de uma abordagem bayesiana.

Em (SAGIROGLU, BAYINDIR *et al.*, 2009), os sistemas de potência ativo têm sido utilizados nos sistemas de energias de forma intensiva. A seleção de um filtro apropriado é muito importante para resolver os problemas de QEE. A seleção de um filtro depende de uma série de parâmetros de sistema de energia, sistemas de apoio à decisão podem ser empregados na seleção de filtro. No trabalho o autor desenvolveu, "Um Sistema de Apoio à Decisão Inteligente, IDSS" é proposto um sistema que determina a eficiência do filtro *shunt* ativo de três fases para os problemas de qualidade de energia. O sistema desenvolvido converte os parâmetros obtidos para os dados úteis no mecanismo de inferência baseado em regras, em seguida, os dados úteis são avaliados em um classificador *Naive Bayes*. Assim, um sistema de apoio à decisão de baixo custo foi desenvolvido para orientar os usuários na taxa de adequação de um sistema trifásico de potência ativa para os problemas de QEE.

Em (TAMANO e TSUJI, 2011), cita-se que precisa-se de métodos eficientes para economizar energia doméstica. Os sistemas domésticos de gerenciamento de energia (HEMS) estão sendo desenvolvidos, mas eles precisam de tempo para mudar o estilo de consumo das pessoas a partir da análise de seu comportamento. Analisou-se um conjunto de dados com a quantidade de consumo e a informação dos consumidores, para apontar que tipo de características dos consumidores, as quais chamou-se de propriedades estáticas, afetariam o consumo. Usou-se a metodologia de aprendizagem de máquinas. Fez-se um classificador *Naive Bayes* para dizer a tendência do consumo das propriedades estáticas do consumidor. Depois de obter a precisão de 0,4148, não se estimou a importância de cada propriedade estática com medidas estatísticas, como x^2 e assim por diante, para melhorar a precisão e encontrar as propriedades estáticas importantes. Embora este processo não leve a uma melhoria significativa da precisão, foram encontradas diversas propriedades estáticas para afetar o consumo. Pode-se adicionar uma funcionalidade de diagnóstico rápido ao HEMS com esses resultados.

Para (MONTENEGRO; RAMOS, 2012), a indústria dos sistemas de potência há um interesse crescente em medir e melhorar a confiabilidade e segurança da rede de energia elétrica de distribuição. No processo de QEE, os distúrbios tornaram-se uma grande preocupação devido à incerteza associada a possíveis problemas causados por esses tipos de distúrbios para as cargas de usuário e operação do utilitário. Como consequência, os métodos probabilísticos e gráficos têm de ser implementados a fim de localizar a causa mais provável e efeito de certos distúrbios de QEE ajudando o analista no processo de tomada de decisão. Redes Bayesianas

são ferramentas úteis para incluir o modelo de conhecimento de falhas em algoritmos de inferência probabilística para determinar as causas mais prováveis e efeitos de um distúrbio QEE em uma situação modelada. Este trabalho apresenta o desenvolvimento de uma ferramenta para diagnóstico de QEE implementadas com LabVIEW®. A ferramenta é aplicada para diagnóstico de SAG em um determinado cenário; os resultados revelam que a ferramenta pode diagnosticar o distúrbio com alta precisão em cada simulação, e representa um bom complemento para os instrumentos de medição de QEE.

O trabalho desenvolvido por (PYLARINOS, THEOFILATOS *et al.*, 2012) faz monitoramento de corrente de vazamento (LC), uma ferramenta amplamente utilizada para a investigação de atividade elétrica de superfície e o desempenho de isoladores de alta tensão. A atividade da superfície está correlacionada com a forma das formas de onda LC. Embora o monitoramento de campo seja necessário para adquirir uma visão exata da atividade e do desempenho dos isoladores, as formas de onda de campo geralmente não são registradas devido ao monitoramento de longo prazo e ao acúmulo de dados.

Em vez disso, os valores extraídos, como o valor do pico, a carga e o número de pulsos que excedem os limiares predefinidos, são gravados, com as formas de onda atuais sendo gravadas ocasionalmente ou não. Contudo, um valor extraído totalmente representativo ainda não foi determinado. A quantidade de 1540 formas de ondas de campo são investigadas para adquirir uma imagem detalhada da forma das formas de onda no campo. As regras de classificação simples são usadas para distinguir entre grupos básicos. As formas de onda de descarga são ainda classificadas com base na duração das descargas. Vinte recursos diferentes, do domínio do tempo e da frequência, dois algoritmos de extração de recursos (*Student t-test e Minimum Redudancy–Maximum Relevance* – ou mRMR) e três algoritmos de classificação (KNN, *Naive Bayes*, SVM) são empregados para a classificação. Os resultados obtidos podem ser utilizados para maximizar a eficiência do monitoramento de campo LC.

Para (SOMASUNDARAM, SIVALINGAM *et al.*, 2013), os algoritmos de reconhecimento de objetos geralmente se concentram na determinação da classe de um objeto detectado em uma cena. Duas fases significativas geralmente estão envolvidas no reconhecimento de objetos. A primeira fase é a fase de representação do objeto, na qual os recursos mais adequados que proporcionam o melhor poder discriminativo sob restrições, como iluminação, resolução, escala e exibição, são descritos para descrever os objetos. A segunda fase é usar este espaço de representação para desenvolver modelos para cada objeto de classe usando classificadores discriminantes. Concentrou-se em objetos compostos, ou seja, objetos

com duas ou mais classes mais simples que estão interligadas de forma complicada. Um exemplo clássico de tal cenário é um ciclista. Um ciclista consiste em uma bicicleta e um humano que anda pela bicicleta. Quando somos confrontados com a tarefa de classificar ciclistas e pedestres, é contra intuitivo e, muitas vezes, difícil encontrar um classificador discriminativo para distinguir as duas classes. Exploramos análise de imagem global com base na representação visual dos resultados com análise de imagem local, na qual se tentou distinguir as partes individuais do objeto composto. Propõe-se uma estrutura unificada *Naive Bayes* e um método de recurso de histograma combinado para combinar os classificadores individuais para um desempenho aprimorado.

Silva *et al.*, (2014) analisa os impactos harmônicos na rede elétrica de uma indústria do Polo Industrial de Manaus (PIM), durante o processo de fabricação de produtos eletrônicos, sendo sua principal fonte de carga não linear: tablets, laptops e computadores pessoais. Todos os dados foram coletados usando analisadores de QEE colocados no fornecimento a partir dos 13,8KV e parâmetro de carga média de 800KVA do processo de fabricação da instalação onde as medidas foram tomadas. O principal objetivo deste trabalho é apresentar uma análise utilizando a técnica de mineração de dados usando classificador (Ingênuo de Bayes), para determinar e quantificar a distorção harmônica nas linhas de produção deste tipo de indústria, utilizando medições simultâneas de tensões e correntes que avaliam como, cada processo afeta a distorção harmônica total da indústria no (PAC) com a energia da concessionária.

O trabalho desenvolvido por (DOORSAMY e CRONJE, 2014), apresenta uma estrutura para a otimização do monitoramento de condição com base na tensão do eixo em geradores síncronos utilizando classificação bayesiana. Com máquinas envolvidas em processos críticos, tais como a geração de energia, é preferível determinar as falhas com antecedência. O sistema proposto utiliza sinais de eixo de tensão como um método on-line para o diagnóstico de falhas incipientes em máquinas síncronas. Um classificador *Naive Bayes* é usado em conjunto com a estimativa do espectro de frequências, a fim de otimizar a técnica de monitoramento de condição da tensão do eixo. Um modelo de Elementos Finitos (FE) e uma máquina experimental são usados para treinar, testar e validar o sistema de classificação de falhas.

O método de avaliação de QEE transitória proposto por (HUANG, GAO *et al.*, 2014), usa o método de classificação *Naive Bayes*, que é baseado em uma grande arquitetura de processamento de dados. Nesta arquitetura, as fontes de dados serão associadas aos aspectos dos dados de monitoramento da rede elétrica, os dados do poder do cliente e os dados públicos,

e a gravidade é classificada em estado normal, estado anormal, estado crítico, de acordo com os resultados da classificação *Naive Bayes*. Com base no tipo de dados de avaliação de QEE transitória, o processamento de dados é usado para poder processar dados distribuídos e dados de transmissão, de modo que ele não só atualize regularmente as regras do classificador como também a avaliação da condição em tempo real. Na fase de treinamento do classificador, usaram-se os dados históricos maciços como o objeto de aprendizagem distribuído afim de gerar regras de avaliação periodicamente. Na fase de avaliação do estado, cada avaliação não atualizará as regras de avaliação geradas pela fase de treinamento, gerará avaliação em tempo real de amostras do trabalho de processamento de fluxo e avaliará o estado de QEE de acordo com a regra atual.

No trabalho desenvolvido por Liu *et. al.* (2015), criou-se um método que tem como objetivo a combinação para a classificação de distúrbios complexos de QEE com base no conjunto de decomposição empírica modo (EEMD) e aprendizagem múltiplos níveis. EEMD é adotado para extrair as características de distúrbios complexos, o que é mais adequado para o processamento de sinal não-estacionário. A função de wavelet, SVM (rank-WSVM) é proposta para ser aplicada na classificação dos distúrbios complexos. Em primeiro lugar, as quantidades características de distúrbios complexos são obtidas mediante a definição de EEMD com diferenças de energia padrão de cada função de modo intrínseco. Em segundo lugar, após a otimização do rank-SVM, com base na função wavelet kernel, a função de ranking e função multi níveis são respectivamente construídas. Por último, rank-WSVM é aplicada para classificar os distúrbios complexos.

No trabalho de Nogueira, (2015), o mesmo aborda a Qualidade da Energia Elétrica (QEE) como constantemente alvo de vários estudos, principalmente os que estão relacionados com o setor de produção industrial, onde estão concentradas grandes cargas dos sistemas elétricos. Com a evolução dos processos de produção industrial e a inserção de novas tecnologias no setor industrial, foram adicionadas grandes quantidades de equipamentos eletrônicos que são fontes geradoras de perturbações nos sistemas, e que alteram a qualidade do produto “energia elétrica”.

A qualidade da energia dos sistemas de energia é afetada por harmônicos extensivos em um sistema de energia interconectado. Torna-se essencial estimar os harmônicos gerados por cargas não lineares. Cálculos deste tipo de harmônico exigem adequadas técnicas de estimativa do estado (Harmonic State Estimation - HSE). A Análise Independente de Componentes é um método eficiente para HSE, pois emprega menos informações sobre a topologia do sistema de

energia. Vários algoritmos (Independent Component Analysis ICA) são usados para HSE, dos quais o COMBI (Combinação de WASOBI (weight adjusted variant of SOBI) e o EFICA “Efficient Variant Fast ICA”), Quando aplicados a um modelo de quatro barras simples, o COMBI é mais preciso do que o EFICA em termos de certos índices de desempenho. A investigação gráfica dos dois algoritmos com a atual corrente harmônica também aponta para a eficiência do algoritmo COMBI sobre EFICA. Propõe-se implementar o problema HSE em um emulador de rede inteligente disponível no laboratório (SUPRIYA e NAMBIAR, 2015).

Para (AZIM, SUN *et al.*, 2015) uma técnica passiva de detecção de ilhas para gerações distribuídas em microgrid, conectadas à rede e apresenta uma análise comparativa abrangente de classificadores inteligentes (Árvore de Decisão, Naive Bayes, Redes Neurais e Máquinas de Suporte a Vetor) para aplicação de detecção de ilhas passivas. O método proposto utiliza técnicas de reconhecimento de padrões na classificação de assinaturas subjacentes de grande variedade de eventos do sistema em parâmetros críticos do sistema para detecção de ilhas. O estudo de caso em um modelo de microgrid conectado à rede com diferentes tipos de gerações distribuídas é realizado para avaliar o método proposto e comparar os desempenhos dos classificadores. Os resultados dos testes demonstram a eficácia do método proposto na detecção de eventos insulares.

Em (KHAN, PADHI *et al.*, 2015) propõe-se uma rápida transformação Time-Time para reconhecimento de padrões e análise de sinal não estacionário responsável pela distorção de QEE. A transformação Time-Time é generalizada e um algoritmo mais rápido é concebido para uma implementação mais rápida. Para distinguir entre vários distúrbios de energia, os recursos estatísticos foram extraídos da distribuição Time-Time. Esses recursos estatísticos são então passados através de um classificador eficiente, nomeadamente o classificador Naive Bayes para identificação de padrões de sinal. O desempenho melhorado pode ser claramente demonstrado através de uma comparação de precisão de classificação entre técnicas propostas anteriormente.

A energia é um recurso limitado, que deve ser gerenciado com sabedoria, levando em consideração tanto a oferta-demanda quanto as restrições de capacidade na rede de distribuição. Um aspecto da gestão de energia inteligente no nível de construção é dado pelo problema da detecção em tempo real da demanda flexível disponível. Foi proposto o uso de técnicas de desagregação de energia para realizar esta tarefa. Em primeiro lugar, investigamos o uso de métodos de classificação existentes para realizar a desagregação energética. Uma comparação é realizada entre quatro classificadores, nomeadamente *Naive Bayes*, *k-Nearest Neighbors*, *Máquina de Suporte a Vetor* e *AdaBoost*. Em segundo lugar, propomos o uso da *Máquina*

Boltzmann Restrita para executar automaticamente a extração de recursos. Os recursos extraídos são então usados como entradas para os quatro classificadores e, conseqüentemente, mostram que eles melhoram sua precisão. A eficiência da nossa abordagem é demonstrada em um banco de dados real consistindo em medidas detalhadas do nível do aparelho com alta resolução temporal, que tem sido utilizada para desagregação de energia em estudos anteriores, a saber, o REDD (*Reference Energy Disaggregation Dataset* - REDD). Os resultados mostram robustez e boas capacidades de generalização para edifícios recém-apresentados com pelo menos 96% de precisão (MOCANU, NGUYEN *et al.*, 2016).

A restauração rápida e eficaz do sistema de energia é essencial para a minimização do tempo de inatividade e do custo que pode se tornar substancialmente rápido após um apagão do sistema. A maioria das redes não possui os sensores para diagnosticar falhas para algoritmos que utilizam essas medidas. Em vez disso, a avaliação depende das chamadas de clientes que perderam energia enquanto a entrada do técnico de campo não é refletida na avaliação de forma formal. Foi proposto investigar a detecção de localização de falhas para restauração de serviços com base em um Sistema de Automação de Distribuição (DAS) com uma abordagem probabilística bayesiana de inteligência centralizada que leva em consideração a opinião subjetiva do técnico de campo. A abordagem emprega lógica probabilística e lógica subjetiva baseada na teoria da evidência. Um modelo simplificado de sistema de distribuição de energia é empregado para introduzir novos conceitos que empregam a teoria da evidência com lógica subjetiva e probabilística para abordar a informação insuficiente (UZUNOGLU, 2017).

Existe uma tendência mundial para a integração das energias renováveis sob a forma de geração distribuída, levando à formação de microgrids. A conexão dessas fontes introduz novos problemas na operação e gerenciamento de sistemas de distribuição. Uma questão importante é a da ilha, onde a micrografia permanece energizada localmente enquanto está isolada da rede principal. É importante detectar este evento Ilha com rapidez e precisão para evitar possíveis danos à (GD geração distribuída) e assegurar que permaneça conectado ao lado GD após a ilha. Foi proposto explorar o problema da detecção oportuna de ilhas por técnicas de aprendizado de máquina, especificamente, consideramos cinco algoritmos, a saber Vizinhos mais Próximos (KNN), classificador de *Bagging*, *Lazy-K **, *Naive Bayes*, e *Random Forest*). Vários casos de ilhas e não ilhas são simulados no sistema de distribuição de 13 barras do IEEE. Diferentes tipos de GDs estão conectados ao sistema e os distúrbios são introduzido. Consideramos o uso da tensão, frequência e sua taxa de variação no ponto de acoplamento comum (PCC) como recursos para a detecção de eventos usando classificadores. Esses recursos

são extraídos dos resultados da simulação e usados para treinar e testar vários tipos de classificadores. É mostrado que o classificador de floresta aleatória detecta a ilha com um alto nível de precisão e dentro de um período de tempo razoável após a ocorrência do distúrbio (AL-BATAINEH e KAVASSERI, 2017).

A complexidade inerente as redes de malha sem fio (*WMNs Wireless Mesh Networks*) dificulta as tarefas de gerenciamento e configuração, especialmente para detecção e diagnóstico de falhas. Além disso, as inspeções manuais são extremamente dispendiosas e exigem uma força de trabalho altamente qualificada, tornando-se assim impraticável à medida que o problema aumenta. Para abordar esta questão, o artigo propõe uma solução que utiliza técnicas de aprendizado de máquinas para detecção e diagnóstico automatizado de falhas (*FDD Fault Detection And Diagnosis*) em redes sem fio (*WMNs*) alimentadas por energia solar. Utilizou-se a metodologia Descoberta de Conhecimento em Banco de Dados (*KDD*) e um dicionário pré-definido de falhas com base em experiência anterior com a implantação de *WMNs*. Posteriormente, o problema foi resolvido como um problema de classificação padrão. Vários algoritmos de classificação foram avaliados, como *Naive Bayes*, *SVM*, Tabela de Decisão, *KNN* e *C.45*. O *SVM* apresentou os melhores resultados, alcançando uma precisão geral de 90,59% durante o treinamento e mais de 85% nos testes de validação (FERREIRA, CARRANO *et al.*, 2017).

As linhas de transmissão são componentes muito importantes do sistema de energia elétrica. Portanto, é necessário prever e detectar tipos de falhas de linhas de transmissão e locais para melhorar o esquema de proteção do sistema de energia e aumentar sua confiabilidade. O artigo investiga o uso de quatro poderosos classificadores de aprendizado de máquina para detectar e prever tipos de falhas e locais em uma linha de transmissão de energia de 750KV, 600km de comprimento. *Bagging*, *Boosting*, *Funções de Base Radial* e classificador *Naive Bayes* foram utilizados para localizar e detectar falhas em uma linha de transmissão de energia. As descobertas mostram que o uso da técnica de aprendizado da máquina pode ser viável para essa tarefa e pode representar uma ótima oportunidade para aumentar a proteção e a eficiência do sistema de energia (HASAN, EBOULE *et al.*, 2017).

A produção industrial evoluiu significativamente na última década. Por esse motivo, é necessário obter ferramentas matemáticas e computacionais que permitam aos engenheiros de sistemas de energia tomar decisões que reduzam as distorções harmônicas de acordo com os padrões internacionais. O artigo apresenta uma avaliação de distorção harmônica total (THD) com base na descoberta de conhecimento completo em bancos de dados (*KDD*) usando padrões

de QEE e ferramentas de inteligência computacional. Os materiais e métodos de avaliação do THD consistem em análise de carga e layout; escolha e instalação de analisadores QEE; e a aplicação do processo completo de KDD, incluindo a coleta, seleção, limpeza, integração, transformação e redução, mineração, interpretação e avaliação dos dados. Esta metodologia de pesquisa foi utilizada em uma indústria elétrica e eletrônica; os resultados obtidos possuem características que podem ser utilizadas como referência para outros tipos de análises. Os resultados indicam que estes métodos podem ser aplicados a várias aplicações industriais, tais como a descrição do processo KDD completo para avaliação do THD no ponto de acoplamento comum (PCC); coleta simultânea usando cinco analisadores QEE em vários pontos da rede elétrica; e o uso de um classificador de árvore de decisão (OLIVEIRA, TOSTES *et al.*, 2017).

Com o intuito de minimizar os transtornos decorrentes da baixa qualidade da energia e os prejuízos causados para as concessionárias e consumidores (industriais, residenciais e comerciais), foi desenvolvido no Brasil, os Procedimentos de Distribuição do Sistema Elétrico Nacional (ANEEL-PRODIST, 2017), criado e desenvolvido pela Agencia Nacional de Energia Elétrica (ANEEL).

O PRODIST tem o objetivo de normatizar e padronizar as atividades relacionadas à distribuição de energia, incluindo os padrões de qualidade do produto. O presente trabalho foi concentrado e realizado em uma Universidade, que possui um sistema elétrico 13.8 KVA, com objetivo de monitorar a qualidade do produto “energia elétrica” através do conteúdo harmônico gerado pela rede elétrica. Os dados gerados foram submetidos a uma técnica de inteligência computacional (IC), com uso do processo de descoberta de extração de conhecimento em bases de dados ou KDD.

O objetivo do trabalho é analisar, identificar e diagnosticar os conteúdos harmônicos representativos para o sistema, podendo assim verificar o quanto cada setor analisado estar afetando a QEE dentro da Universidade e no ponto de acoplamento com a concessionária, através das gerações de distorções harmônicas, evitando assim, penalizações e outras sanções regulamentadas.

2.3 Considerações Finais

Neste capítulo, apresentou-se inúmeros trabalhos que possuem uma correlação com as áreas de energia elétrica, QEE, eletrônica de potência, inteligência computacional e mineração de dados. Demonstrando a aplicação de técnicas de inteligência computacional, mineração de dados e

KDD nas mais diversas áreas do conhecimento com a mesma eficácia. Os trabalhos apresentados no presente capítulo utilizam uma variedade de técnicas de inteligência computacional, presente no KDD as quais serão utilizadas como base para o desenvolvimento dos capítulos a seguir deste trabalho.

CAPÍTULO III

QUALIDADE DE ENERGIA ELÉTRICA E NORMAS

3.1 Introdução

A qualidade do fornecimento de energia elétrica tem ocupado cada vez mais importância tanto para as distribuidoras quanto para os consumidores desta energia. Neste capítulo discutiremos sobre QEE e normas técnicas nacionais e internacionais que estabelecem os limites de qualidade que o sistema elétrico deve cumprir para oferecer um bom serviço elétrico aos consumidores.

Atualmente o sistema elétrico vem experimentando um aumento da sua capacidade de geração, transmissão e distribuição. Os avanços da tecnologia ocorridos nas últimas décadas propiciaram a obtenção de grandes desenvolvimentos na área da eletrônica, o aumento da utilização de cargas não lineares por consumidores residenciais, comerciais e industriais, tem exigido das concessionárias de energia uma preocupação cada vez maior em relação à QEE nos sistemas elétricos de potência. Tais cargas conhecidas como “Não-lineares” ou “Cargas Elétricas Especiais” provocam distorções de tensão e/ou corrente nas redes elétricas.

Existem normas internacionais e nacionais para o controle da distorção das formas de onda de tensão e corrente. Entre elas destacam-se: a IEEE Std. 519TM – 2014 (Revisão da IEEE Std 519-1992) e a IEC 61000-3-6 (IEC/TR, 2008-02), IEC 61000-4-7 Ed. 2.1 (2009-10), IEC 61000-4-30 Ed. 2.0 (2008-10). No Brasil, a Agência Nacional de Energia Elétrica – ANEEL, dispõe dos Procedimentos de Distribuição de Energia Elétrica do Sistema Elétrico Nacional – PRODIST – Módulo 8 - Revisão 10, Resolução Normativa nº 794/2017, data de vigência 01/01/2018.

3.2 Conceitos Gerais de Qualidade de Energia

A QEE é a condição do sinal elétrico de tensão e corrente que permitem os equipamentos, processos, instalações e sistemas elétricos operarem de forma satisfatória, sem prejuízo de desempenho e de vida útil. A expressão é usada para descrever a energia elétrica que aciona uma carga elétrica, fazendo-a operar corretamente. Sem uma alimentação adequada,

com certo padrão de qualidade, a carga pode operar inadequadamente ou incorretamente, falhar prematuramente ou simplesmente não funcionar (LEÃO, ANTUNES *et al.*, 2015).

As concessionárias de energia elétrica e usuários finais de energia elétrica estão se tornando cada vez mais preocupados com a qualidade da energia elétrica. O termo qualidade de energia tornou-se um dos chavões mais prolíficos da indústria de energia desde o final da década de 1980. É um conceito genérico para uma infinidade de diversos tipos de distúrbios do sistema de potência. As questões que recaem sob esse assunto não são necessariamente novos. O que é novo, é que os engenheiros estão agora tentando lidar com estas questões, utilizando uma abordagem do sistema, em vez de lidar com eles como problemas individuais (DUGAN *et al.*, 2004).

DUGAN *et al.*, (2004), define qualidade de energia como sendo qualquer problema manifestado no desvio da corrente, tensão ou de frequência que resulte em falhas ou falhas de operação dos equipamentos dos clientes. Não pode haver definições completamente diferentes para qualidade de energia, dependendo de um quadro de referência. Por exemplo, um utilitário pode definir a qualidade de energia com confiabilidade e demonstrando estatísticas de que o seu sistema é 99,98 por cento confiáveis. Os Critérios estabelecidos pelas agências reguladoras são geralmente nesse sentido. O fabricante de equipamentos de carga pode definir a qualidade de energia como as características do fornecimento de energia que permitem que o equipamento funcione adequadamente. Essas características podem ser muito diferentes para diferentes critérios.

Para Sankaran (2001), uma definição mais simples e talvez mais concisa poderia dizer: "A qualidade de energia é um conjunto de limites elétricos que permitem que um equipamento funcione da maneira pretendida sem perda significativa de desempenho ou expectativa de vida". Esta definição engloba dois aspectos que exigimos de um dispositivo elétrico: desempenho e expectativa de vida. Qualquer problema de energia que comprometa qualquer atributo é uma preocupação de qualidade de energia.

Para Tostes (2013), ainda não existe consenso para o termo "Qualidade da Energia Elétrica", podendo ter significados diferentes para pessoas distintas. Uma definição aceita é: Energia elétrica de boa qualidade é aquela que garante o funcionamento contínuo, adequado e seguro dos equipamentos elétricos e processos associados, sem afetar o meio ambiente e o bem-estar das pessoas.

3.3 Distorções Harmônicas

Harmônicos são correntes ou tensões senoidais de frequências múltiplas (de inteiros) da frequência que o sistema é designado a operar. Os componentes harmônicos, combinados com a tensão ou corrente fundamentais, produzem alterações na forma de onda. A distorção harmônica existe devido a características não lineares de dispositivos e cargas do sistema elétrico. A distorção de tensão resulta da queda de tensão provocada pela passagem de corrente (injetada por uma carga não linear) pela impedância do sistema.

Para Arrilaga (2003) é importante ressaltar que a distorção harmônica é um fenômeno que deve ser tratado como sendo de regime permanente. A distorção de forma de onda, provocada pelos componentes harmônicos, deve estar presente, continuamente, por pelo menos alguns segundos.

Segundo Dugan *et al* (2003), com a presença dos harmônicos nos sistemas elétricos, os mesmos passam a não operar mais em condições senoidais, logo as simplificações realizadas por engenheiros nas análises para a frequência fundamental não são mais aplicáveis.

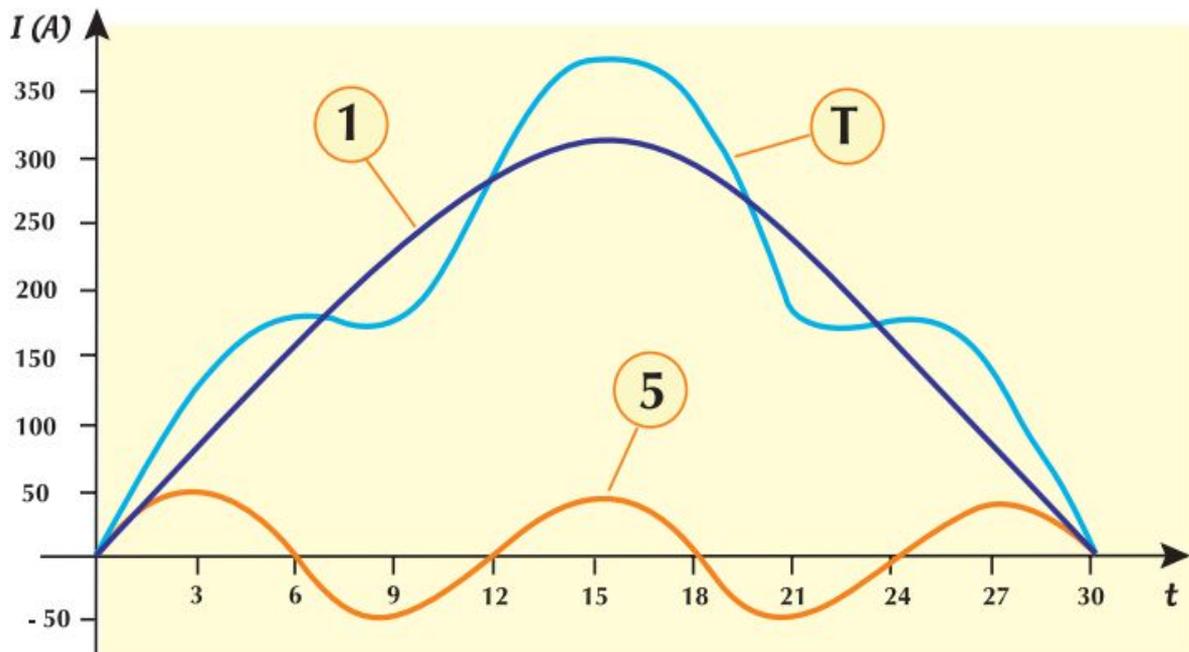
A presença de harmônicos num sistema elétrico provoca distorções na forma de onda da tensão e da corrente, uma vez que os mesmos se somam à componente fundamental. Além disso, os harmônicos são gerados por equipamentos ou cargas que apresentam características não lineares entre tensão e corrente.

Para o IEEE 519-2014, aplicam-se os seguintes termos e definições. O componente de 3ª harmônico: É um componente de ordem maior que a fundamental em uma série de Fourier. Por exemplo, em um sistema de 60 Hz, a harmônica de 3ª. ordem, também conhecida como "terceira harmônica", é 180 Hz.

Para (MORENO, 2001) os harmônicos são uma tensão ou corrente pode ser definida como um sinal senoidal cuja frequência é um múltiplo inteiro da frequência fundamental do sinal de alimentação. Desta forma, pode-se dizer que um sinal periódico contém harmônicos quando a forma de onda desse sinal não é senoidal ou, dito de outro modo, um sinal contém harmônicos quando ele é deformado em relação a um sinal senoidal.

Em (MORENO, 2001) a forma de onda de tensão ou de corrente em um dado ponto de uma instalação pode ter o aspecto do sinal T que está mostrado na figura 1 (onda deformada). Observando essa situação, percebe-se que o sinal T é a soma ponto a ponto dos sinais 1 e 5 formados por senóides perfeitas de amplitudes e frequências diferentes, chamadas de harmônicas.

Figura 1 – Onda deformada e suas componentes harmônicas



Fonte Moreno (2001).

Com a proliferação de cargas não-lineares, os harmônicos vêm ganhando maior destaque em estudos de qualidade da energia elétrica, a fim de evitar, ou pelo menos reduzir, as indesejadas consequências provenientes destes nos diversos elementos do sistema elétrico. Além disso, estudos relacionados à contribuição de cada carga ou conjunto destas na distorção harmônica de uma determinada barra do sistema vem sendo desenvolvidos a fim de atribuir responsabilidades aos consumidores conectados à rede elétrica.

As fontes harmônicas de baixa potência, como PCs e TVs, lâmpadas fluorescentes com reatores eletrônicos, pelo seu grande número nas instalações comerciais, de ensino e residenciais, etc., vieram a constituir-se, também, em fontes disseminadoras de harmônicos. Além disso, muitas destas cargas poluidoras são, também, sensíveis às distorções harmônicas. Os harmônicos de corrente são produzidos por cargas não lineares, tais como equipamentos de eletrônica de potência. Essas cargas geram correntes não senoidais mesmo sendo alimentadas com tensão senoidal. Essas correntes distorcidas ao circularem pela impedância do sistema, que é constituída pela impedância da fonte mais impedância da fiação e transformadores, provocam a distorção da onda de tensão. Essa é a origem dos harmônicos de tensão, pois a concessionária produz, na geração, uma onda senoidal pura de tensão.

Os harmônicos que causam maior impacto costumam ser de ordem ímpar, ou seja, 3^a, 5^a, 7^a, 9^a, etc., porque as ondas de tensão ou corrente costumam ter simetria ímpar e, portanto, o

semiciclo positivo da onda tem simetria com o semiciclo negativo da onda. Como na prática não existe uma simetria exata, ao se analisar o espectro de uma forma de onda real, aparecem algumas componentes de ordem par de amplitude pequena. Para fins didáticos considera-se que não existem harmônicos de ordem par para a maioria das cargas. Uma exceção é o caso dos fornos a arco em que as componentes harmônicas de ordem par são relevantes.

3.3.1 Ordem, Frequência e Sequências das Harmônicas

Os sinais harmônicos são classificados quanto a sua ordem, frequência e sequência conforme tabela 1 a seguir. Em se tratando de sistemas elétricos trifásicos equilibrados, quando a decomposição de uma forma de onda de tensão ou corrente distorcida, as ordens harmônicas advindas desta decomposição podem ser estabelecidas em termos de componentes simétricos. Em uma situação ideal, onde somente existisse um sinal de frequência 60 Hz, apenas existiria a harmônica de ordem 1, chamada de fundamental.

Observando a tabela 1 a seguir, percebe-se dois tipos de harmônica: as ímpares e pares. As ímpares são encontradas nas instalações elétricas em geral e as harmônicas pares existem nos casos de haver assimetrias do sinal devido a presença de componentes contínuas. A sequência pode ser positiva, negativa ou nula (zero). Tomando-se como exemplo um motor assíncrono trifásico alimentado por quatro condutores (3F+ N), os harmônicos de sequência positiva tenderiam a fazer o motor girar no mesmo sentido que o da componente fundamental, provocando assim, uma sobre corrente nos seus enrolamentos, que provocaria um aumento de temperatura, reduzindo a vida útil e permitindo a ocorrência de danos ao motor. Essas harmônicas de sequência positiva provocam, geralmente, aquecimentos indesejados em condutores, motores, transformadores, etc. (MORENO, 2001).

Tabela 1 - Ordem, frequência e sequência das harmônicas.

Ordem	Frequência (Hz)	Sequência
1	60	+
2	120	-
3	180	0
4	240	+
5	300	-
6	360	0
<i>n</i>	<i>n.60</i>	-

Fonte: Moreno, (2001).

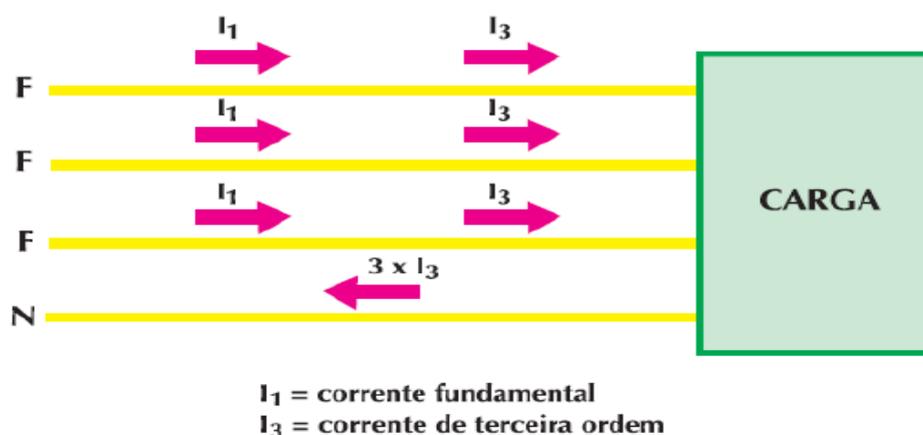
Em sistemas significativamente desequilibrados, cada harmônico pode ser decomposto nas três componentes simétricas, ou seja, sequência positiva, sequência negativa e sequência zero.

Os harmônicos de sequência negativa fariam o motor girar em sentido contrário ao giro produzido pela fundamental, freando assim o motor e também causando aquecimento indesejado. Por sua vez, as harmônicas de sequência nula, zero ou também conhecidas como homopolares, não provocam efeitos no sentido de rotação do motor, porém somasse as correntes algebricamente no condutor neutro, isso implica que podem ocorrer situações em que pelo condutor neutro pode circular uma corrente de terceira ordem que é três vezes maior que a corrente de terceira ordem que percorre cada condutor fase. Com isso, ocorrem aquecimento excessivo do condutor neutro, destruição de bancos de capacitores, etc. (MORENO, 2001).

A natureza e a intensidade dos harmônicos gerados por cargas não lineares dependem de cada carga especificamente, mas três aspectos devem ser considerados gerais: – É um fenômeno contínuo, ou seja, de longa duração; – Devido ao comportamento similar no semiciclo positivo e no semiciclo negativo da quase totalidade das cargas, os harmônicos de ordem ímpar são mais frequentes e com maior intensidade e, por isso, geralmente são os causadores de problemas; – Quanto maior a ordem ou frequência do harmônico menor sua intensidade. Conhecidos os valores de tensões ou correntes harmônicas presentes no sistema, utilizam-se de procedimentos quantitativos para expressar a influência do conteúdo harmônico em uma forma de onda.

Um dos mais utilizados é a “Distorção Harmônica Total”, a qual pode ser empregada tanto para sinais de tensões como para correntes, a figura 2 mostra a fundamental e o terceiro harmônico.

Figura 2 - Harmônico de sequência zero.



Fonte: Moreno, (2001).

Normalmente, quando a análise de componentes harmônicos de tensão e/ou corrente é realizada sobre os sistemas de potência, não leva-se em consideração a extração das frequências harmônicas acima da 63ª ordem, pois os equipamentos necessitariam de maior precisão, a qual viria acompanhada por um alto custo associado ao equipamento de inferência. (DUGAN *et al.*, 2001).

3.3.2 Cargas não-lineares

Carga não-linear é aquela que, quando alimentada com tensão senoidal, absorve uma corrente não senoidal, ou seja, a carga não linear tem a característica de deformar a forma de onda da corrente absorvida. Nesta definição está implícito que se trata de um regime permanente senoidal, por isso, se diz que uma carga linear é aquela onde existe uma proporcionalidade entre tensão e corrente através de uma impedância Z a qual afeta a amplitude e a fase da corrente mantendo a forma de onda senoidal; o mesmo não acontece com as cargas não lineares que deformam a forma de onda e por isso também são chamadas de cargas deformantes (DIAS, 1998).

Cargas não lineares causam correntes distorcidas mesmo quando alimentadas por uma fonte com tensão não distorcida (senoidal). Se uma tensão distorcida é aplicada sobre uma carga linear, a forma de onda da corrente na carga será distorcida como a tensão. Entretanto, se uma carga não linear é alimentada por uma tensão não senoidal, a forma de onda da corrente será distorcida por causa da tensão distorcida e da não linearidade da carga. Em quaisquer desses casos, a relação entre a tensão e a corrente não é constante (LEÃO, ANTUNES *et al.*, 2015).

Os dispositivos geradores de harmônicas estão presentes em todos os sectores industriais, terciários e domésticos. As harmônicas são o efeito de cargas não lineares. Uma carga é dita não linear quando a corrente que ela absorve não tem a mesma forma que a tensão que a alimenta. Tipicamente, as cargas que utilizam electrónica de potência são não lineares.

Exemplos de cargas não lineares: informática, pontes retificadoras, variadores de velocidade, fornos a arco, iluminação fluorescente.

Estas são cargas nas quais a corrente que por ela circula não é diretamente proporcional a tensão fornecida. Por conseguinte, qualquer carga que requisita uma corrente não senoidal de uma tensão senoidal é não linear.

3.4 Normas de Qualidade de Energia

A preocupação com a qualidade da energia elétrica fornecida aos consumidores nasceu junto com as primeiras experiências comerciais relacionadas com a geração, a transmissão e a distribuição de energia, no século XIX. Já em 1934 a legislação brasileira estabelecia, em seu Código de Águas, os primeiros indicadores de controle sobre esta qualidade.

É decorrente em parte da reformulação que o setor elétrico vem experimentando, para viabilizar a implantação de um mercado consumidor, no qual o produto comercializado passa a ser a própria energia elétrica.

As normas de qualidade de energia compreendem um conjunto de regulações que estabelecem os limites de qualidade que o sistema elétrico deve cumprir para oferecer um bom serviço elétrico aos consumidores.

As normas e recomendações internacionais também são usadas como referência para a avaliação da qualidade da energia elétrica. São elas:

Padrões do IEEE mais comuns usados em QE:

- IEEE 446 - *Emergency and Standby Power*
- IEEE 519 - *Harmonic Control*
- IEEE 1001 - *Interface with Dispersed Generation*
- IEEE 1100 - *Power and Grounding Electronics*
- IEEE 1159 - *Monitoring Power Quality*
- IEEE 1250 - *Service to Critical Loads*
- IEEE 1346 - *System Compatibility in Industrial Environments*
- IEEE 1366 - *Electric Utility Reliability Indices*

Padrões do IEC mais comuns usados em QE:

- IEC 61000-2-2 - *General guide on harmonics and interharmonics measurements and instrumentation for power supply systems and equipment connected thereto*
- IEC 61000-2-4 - *Compatibility levels in industrial plants lowfrequency conducted disturbances*
- IEC 61000-4-7 - *Compatibility levels for low-frequency conducted disturbances and signaling public low-voltage power supply systems*
- IEC 61000-4-11 - *Voltage dips/interruptions/variation immunity*
- IEC 61000-4-14 - *Voltage fluctuation immunity*

- IEC 61000-4-13 - *Harmonics/interharmonics immunity*
- IEC 61000-4-15 - *Flickermeter*
- IEC 61000-4-16 - *Test for immunity to conducted common mode disturbances in the frequency range 0 Hz to 150 kHz*
- IEC 61000-4-17 - *Ripple on d.c. input power port, immunity test*
- IEC 61000-4-27 - *Unbalance, immunity test*
- IEC 61000-4-28 - *Variation of power frequency, immunity test*
- IEC 61000-4-29 - *Voltage dips, short interruptions and voltage variations on D.C. input power port, immunity tests*
- IEC 61000-4-30 - *Power quality measurements methods*

No Brasil a ANEEL Agência Nacional de Energia Elétrica, autarquia em regime especial vinculada ao Ministério de Minas e Energia, foi criada para regular o setor elétrico brasileiro, por meio da Lei nº 9.427/1996 e do Decreto nº 2.335/1997.

Os Procedimentos de Distribuição de Energia Elétrica no Sistema Elétrico Nacional – PRODIST, são documentos elaborados pela ANEEL e normatizam e padronizam as atividades técnicas relacionadas ao funcionamento e desempenho dos sistemas de distribuição de energia elétrica.

O PRODIST Módulo 8 trata da Qualidade da Energia Elétrica a versão vigente, revisão 10 tem como data de vigência, o período a partir de 01/01/2018.

O PRODIST estabelecer os procedimentos relativos à qualidade da energia elétrica - QEE, abordando a qualidade do produto e a qualidade do serviço.

Para a qualidade do produto, este módulo define a terminologia e os indicadores, caracteriza os fenômenos, estabelece os limites ou valores de referência, a metodologia de medição, a gestão das reclamações relativas à conformidade de tensão em regime permanente e às perturbações na forma de onda de tensão e os estudos específicos de qualidade da energia elétrica para fins de acesso aos sistemas de distribuição.

3.4.1 Norma IEC 61000-3-6 de Compatibilidade Eletromagnética

A série 61000 da IEC (Comissão Eletrotécnica Internacional) compreende um amplo conjunto de regulações sobre a qualidade de energia. Em particular, a IEC 61000-3-6 (IEC/TR,

61000-3-6, 2008) estabelece os limites de emissão de harmônicos para as instalações conectadas a sistemas elétricos de potência. Para a tensão elétrica esta norma define:

1) A Distorção Harmônica Total de Tensão (%THDV). Calculada como a relação entre a tensão *rms* dos harmônicos superiores considerados e a tensão fundamental (V_1), apresentado em (1).

$$\%THDV = \frac{\sqrt{\sum_{h=2}^{\infty} V_h^2}}{V_1} 100 \quad (1).$$

2) A Distorção Harmônica Individual de tensão de ordem h (%IHDV $_h$). Calculada como a relação entre a tensão de um harmônico individual (V_h) e a tensão fundamental de acordo com (2).

$$\%IHDV_h = \frac{V_h}{V_1} 100 \quad (2).$$

A referência (IEC/TR 61000-3-6, 2008) estabelece dois tipos de limites: (1) níveis de compatibilidade como é mostrado na Tabela 2 e (2) níveis de planejamento, como apresentado na Tabela 3. Quando as emissões de harmônicos das cargas não lineares não superam os níveis de compatibilidade estabelecidos, assegura-se uma boa qualidade de energia. Por outro lado, os níveis de planejamento (mais restritivos que os de compatibilidade) são seguidos como guias para o planejamento do sistema e os mesmos asseguram o cumprimento dos níveis de compatibilidade.

Tabela 2 - Níveis de compatibilidade para as tensões harmônicas individuais em redes de baixa(BT) e média tensão (MT).

Ímpares não múltiplas de 3		Ímpares múltiplas de 3		Pares	
Ordem de Harmônico h	Tensão Harmônica %	Ordem de Harmônico h	Tensão Harmônica %	Ordem de Harmônico h	Tensão Harmônica %
5	6	3	2	2	2
7	5	9	1.5	4	1
11	3.5	15	0.4	6	0.5
13	3	21	0.3	8	0.5
$17 \leq h \leq 49$	$2.27(17/h) - 0.27$	$21 \leq h \leq 45$	0.2	$10 \leq h \leq 50$	$0.25(10/h) + 0.25$

Fonte: (IEC/TR 61000-3-6, 2008).

Os limites de harmônicos de tensão são estabelecidos para evitar os efeitos nocivos dos harmônicos em regime permanente e em curto prazo que se define como:

1) Efeitos de regime permanente são relacionados fundamentalmente a efeitos térmicos em capacitores, cabos, transformadores, motores e outros, e são medidos em média a intervalo de 10 minutos.

2) Efeitos a curto prazo que se manifestam em equipamentos eletrônicos sensíveis a níveis de harmônicos tendo como intervalo de interesse para registro a faixa de 3 segundos ou menos.

O nível de compatibilidade do THDV para média e baixa tensão é de 8% para harmônicos de regime permanente (intervalos de 10 minutos) e de 11% para harmônicos de curta duração (intervalos menores que 3 segundos). Os limites de compatibilidade da Tabela 2 e de planejamento da Tabela 3 para as tensões harmônicas individuais são válidos para harmônicos estacionários. Estes limites devem ser modificados pelo fator K_{hvs} para ser utilizados com harmônicos de curta duração conforme (3).

$$K_{hvs} = 1.3 + 0.7 \frac{h-5}{45} \quad (3)$$

Tabela 3 - Níveis de planejamento para tensões harmônicas individuais em redes de média

Ímpares não múltiplas de 3			Ímpares múltiplas de 3			Pares		
Ordem de Harmônico h	Tensão Harmônica h %		Ordem de Harmônico h	Tensão Harmônica %		Ordem de Harmônico h	Tensão Harmônica %	
	MT	AT-EHT		MT	AT-EHT		MT	AT-EHT
5	5	2	3	4	2	2	1.8	1.4
7	4	2	9	1.2	1	4	1	0.8
11	3	1.5	15	0.3	0.3	6	0.5	0.4
13	2.5	1.5	21	0.2	0.2	8	0.5	0.4
$17 \leq h \leq 49$	$1.9(17/h)-0.2$	$1.2(17/h)$	$21 \leq h \leq 45$	0.2	0.2	$10 \leq h \leq 50$	$0.25(10/h)+0.22$	$0.19(10/h)+0.16$

Fonte: (IEC/TR 61000-3-6, 2008).

A referida norma não estabelece limites explícitos para a distorção da corrente no PCC, no entanto, a mesma contempla a possibilidade de converter os limites de distorção para a tensão em limites de distorção para a corrente ao utilizar a impedância em frequências harmônicas do sistema externo de energia.

3.4.2 Norma IEEE 519-2014 Prática Recomendada IEEE e Requisitos para Controle Harmônicos em Sistemas Elétricos de Potência

A norma IEEE 519-1992 que sofreu alterações e atualmente é chamada IEEE 519-2014. É um documento largamente adotado para controle dos harmônicos no ponto de

acoplamento elétrico entre a indústria e a concessionária. A filosofia desta norma é não se preocupar com o que ocorre no interior da instalação e sim com o que a instalação pode injetar na rede e, portanto, atingir outros consumidores.

A referida norma, propõe um compartilhamento de responsabilidade para a manutenção das harmônicas no PAC, onde os consumidores devem garantir que as correntes harmônicas estejam limitadas e as concessionárias, por sua vez, garantam os limites das harmônicas de tensão. Esta recomendação propõe a avaliação das harmônicas no PAC, pois assim, pode-se determinar como um consumidor afeta tanto a concessionária quanto outro consumidor conectado ao mesmo alimentador.

No PAC, os proprietários ou operadores de sistemas devem limitar harmônicos de tensão de linha a neutro da seguinte maneira:

- Período diário de 99º percentil, valores de tempo muito curto (3s) devem ser inferiores a 1,5 vezes os valores dados na Tabela 4.

- Os valores de percentil 95º, percentil semanal (10 min) devem ser inferiores aos valores indicados na Tabela 4.

Todos os valores devem estar em porcentagem da tensão de frequência de potência nominal no PAC. A Tabela 4 aplica-se a harmônicos de tensão cujas frequências são múltiplos inteiros da frequência de energia.

Todos os valores devem estar em percentual da tensão de frequência de potência nominal no PAC. Tabela 4 aplica-se a tensão harmônicos cujas frequências são múltiplos inteiros da frequência de alimentação.

Tabela 4 - Limites de distorção de tensão.

Tensão do barramento V no PCC	Harmônico individual (%)	Distorção harmônica total THD (%)
$V \leq 1.0 \text{ kV}$	5.0	8.0
$1 \text{ kV} < V \leq 69 \text{ kV}$	3.0	5.0
$69 \text{ kV} < V \leq 161 \text{ kV}$	1.5	2.5
$161 \text{ kV} < V$	1.0	1.5 ^a

Fonte IEEE 519-2014, (2014).

Os sistemas de alta tensão podem ter até 2,0% de THD onde a causa é um terminal HVDC cujos efeitos terão atenuado em pontos na rede onde futuros usuários podem estar conectados.

Tabela 5 - Limites de distorção de corrente para sistemas de 120 V a 69 KV.

Distorção máxima da corrente harmônica em porcentagem de I_L						
Ordem harmônica individual (harmônicos ímpares) ^{a, b}						
I_{sc}/I_L	$3 \leq h < 11$	$11 \leq h < 17$	$17 \leq h < 23$	$23 \leq h < 35$	$35 \leq h \leq 50$	TDD
$< 20^c$	4.0	2.0	1.5	0.6	0.3	5.0
$20 < 50$	7.0	3.5	2.5	1.0	0.5	8.0
$50 < 100$	10.0	4.5	4.0	1.5	0.7	12.0
$100 < 1000$	12.0	5.5	5.0	2.0	1.0	15.0
> 1000	15.0	7.0	6.0	2.5	1.4	20.0

Fonte IEEE 519-2014, (2014).

a. As harmônicas estão limitadas a 25% dos limites harmônicos ímpares acima.

b. As distorções atuais que resultam em um deslocamento em dc, por exemplo, conversores de meia-onda, não são permitidas.

b. Todo o equipamento de geração de energia está limitado a esses valores de distorção atual, independentemente da I_{sc} / I_L real.

Onde

I_{sc} = corrente máxima de curto-circuito no PAC

I_L = corrente máxima de carga da demanda (componente de frequência fundamental) no PAC sob condições normais de operação da carga.

Tabela 6 - Limites de distorção de corrente para sistemas com classificação acima de 69 kV até

Distorção máxima da corrente harmônica em porcentagem de I_L						
Ordem harmônica individual (harmônicos ímpares) ^{a, b}						
I_{sc}/I_L	$3 \leq h < 11$	$11 \leq h < 17$	$17 \leq h < 23$	$23 \leq h < 35$	$35 \leq h \leq 50$	TDD
$< 20^c$	2.0	1.0	0.75	0.3	0.15	2.5
$20 < 50$	3.5	1.75	1.25	0.5	0.25	4.0
$50 < 100$	5.0	2.25	2.0	0.75	0.35	6.0
$100 < 1000$	6.0	2.75	2.5	1.0	0.5	7.5
> 1000	7.5	3.5	3.0	1.25	0.7	10.0

Fonte IEEE 519-2014, (2014).

a. Mesmo os harmônicos são limitados a 25% dos limites harmônicos ímpares acima.

b. As distorções de corrente que resultam em um deslocamento em dc, por exemplo, conversores de meia-onda, não são permitidas.

c. Todo o equipamento de geração de energia está limitado a esses valores de distorção atual, independentemente de I_{sc} / I_L real.

Onde

I_{sc} = corrente máxima de curto-circuito no PAC

I_L = corrente de carga de demanda máxima (componente de frequência fundamental) no PAC em condições normais de operação de carga.

Tabela 7 - Limites de distorção de corrente para sistemas classificados > 161 Kv.

Distorção máxima da corrente harmônica em porcentagem de I_L						
Ordem harmônica individual (harmônicos ímpares) ^{a, b}						
I_{sc}/I_L	$3 \leq h < 11$	$11 \leq h < 17$	$17 \leq h < 23$	$23 \leq h < 35$	$35 \leq h \leq 50$	TDD
$< 25^c$	1.0	0.5	0.38	0.15	0.1	1.5
$25 < 50$	2.0	1.0	0.75	0.3	0.15	2.5
≥ 50	3.0	1.5	1.15	0.45	0.22	3.75

Fonte IEEE 519-2014, (2014).

- a. As harmônicas estão limitadas a 25% dos limites harmônicos ímpares acima.
- b. As distorções atuais que resultam em um deslocamento em dc, por exemplo, conversores de meia-onda, não são permitidas.
- c. Todo o equipamento de geração de energia está limitado a esses valores de distorção atual, independentemente da I_{sc} / I_L real.

Onde

I_{sc} = corrente máxima de curto-circuito no PAC

I_L = corrente máxima de carga da demanda (componente de frequência fundamental) no PAC sob condições normais de operação da carga

3.4.3 Norma ANEEL - Procedimentos de Distribuição de Energia Elétrica no Sistema Elétrico Nacional – PRODIST - Módulo 8

O módulo 8 aborda tanto a qualidade do serviço quanto a do produto, neste trabalho estamos interessados na qualidade do produto. Para a qualidade do produto, este módulo define a terminologia e os indicadores, caracteriza os fenômenos, estabelece os limites ou valores de referência, a metodologia de medição, a gestão das reclamações relativas à conformidade de tensão em regime permanente e às perturbações na forma de onda de tensão e os estudos específicos de qualidade da energia elétrica para fins de acesso aos sistemas de distribuição, **PRODIST** (2018).

Os procedimentos de qualidade da energia elétrica definidos neste módulo devem ser observados por:

- a) consumidores com instalações conectadas em qualquer classe de tensão de distribuição;
- b) centrais geradoras;
- c) distribuidoras;
- d) agentes importadores ou exportadores de energia elétrica;
- e) transmissoras detentoras de Demais Instalações de Transmissão – DIT;

f) Operador Nacional do Sistema Elétrico – NOS.

O PRODIST módulo 8 define as terminologias, caracteriza os fenômenos e estabelece os indicadores e limites ou valores de referência relativos à conformidade de tensão em regime permanente e às perturbações na forma de onda de tensão. Aborda os seguintes fenômenos da qualidade do produto em regime permanente:

- i. tensão em regime permanente;
- ii. fator de potência;
- iii. harmônicos;
- iv. desequilíbrio de tensão;
- v. flutuação de tensão;
- vi. variação de frequência.

Existem muitos outros fatores relevantes além da Qualidade do Produto definido e padronizados através do PRODIST Módulo 8 Revisão Dez (10) 2018, porém o aspecto que será tratado com maior relevância no atual trabalho, será as distorções harmônicas, as quais nos servirão como as principais variáveis de análise na tomada de decisão. As distorções harmônicas são fenômenos associados a deformações nas formas de ondas das tensões e correntes em relação à onda senoidal da frequência fundamental, PRODIST (2018).

A tabela 8 mostra as terminologias aplicáveis ao cálculo das distorções harmônicas.

Tabela 8 - Terminologia.

Identificação da Grandeza	Símbolo
Distorção harmônica individual de tensão de ordem h	$DTI_h\%$
Distorção harmônica total de tensão	$DTT\%$
Distorção harmônica total de tensão para as componentes pares não múltiplas de 3	$DTTP\%$
Distorção harmônica total de tensão para as componentes ímpares não múltiplas de 3	$DTTi\%$
Distorção harmônica total de tensão para as componentes múltiplas de 3	$DTT_3\%$
Tensão harmônica de ordem h	V_h
Ordem harmônica	h
Ordem harmônica máxima	h _{máx}
Ordem harmônica mínima	h _{min}
Tensão fundamental medida	V_1
Valor do indicador $DTT\%$ que foi superado em apenas 5 % das 1008 leituras válidas	$DTT_{95\%}$
Valor do indicador $DTTP\%$ que foi superado em apenas 5 % das 1008 leituras válidas	$DTT_{p95\%}$
Valor do indicador $DTTi\%$ que foi superado em apenas 5 % das 1008 leituras válidas	$DTT_i95\%$
Valor do indicador $DTT_3\%$ que foi superado em apenas 5 % das 1008 leituras válidas	$DTT_{395\%}$

Fonte PRODIST, (2018).

Tabela 9 - Expressões para o cálculo das grandezas DIT_h%, DTT%, DTT_p%, DTT_i% e DTT₃% são.

sendo: h = ordem harmônica individual.	$DIT_h\% = \frac{V_h}{V_1} \times 100$
sendo: h = todas as ordens harmônicas de 2 até hmáx. hmáx = conforme a classe A ou S.	$DTT\% = \frac{\sqrt{\sum_{h=2}^{h_{\max}} V_h^2}}{V_1} \times 100$
sendo: h = todas as ordens harmônicas pares, não múltiplas de 3 (h = 2, 4, 8, 10, 14, 16, 20, 22, 26, 28, 32, 34, 38, ...). hp = máxima ordem harmônica par, não múltipla de 3.	$DTT_p\% = \frac{\sqrt{\sum_{h=2}^{h_p} V_h^2}}{V_1} \times 100$
sendo: h = todas as ordens harmônicas ímpares, não múltiplas de 3 (h = 5, 7, 11, 13, 17, 19, 23, 25, 29, 31, 35, 37, ...). hi = máxima ordem harmônica ímpar, não múltipla de 3.	$DTT_i\% = \frac{\sqrt{\sum_{h=5}^{h_i} V_h^2}}{V_1} \times 100$
sendo: h = todas as ordens harmônicas múltiplas de 3 (h = 3, 6, 9, 12, 15, 18, 21, 24, 27, 30, 33, 36, 39, ...). h3 = máxima ordem harmônica múltipla de 3.	$DTT_3\% = \frac{\sqrt{\sum_{h=3}^{h_3} V_h^2}}{V_1} \times 100$

Fonte: Autor, (2019).

Limites para os indicadores.

Tabela 10 - Os limites para as distorções harmônicas totais.

Indicador	Tensão nominal		
	V _n ≤ 1,0 kV	1,0 kV < V _n < 69 kV	69 kV < V _n < 230 kV
DTT ₉₅ %	10,0 %	8,0 %	5,0 %
DTT _{p95} %	2,5 %	2,0 %	1,0 %
DTT _{i95} %	7,5 %	6,0 %	4,0 %
DTT ₃₉₅ %	6,5 %	5,0 %	3,0 %

Fonte: PRODIST, (2018).

Os limites correspondem ao máximo valor desejável a ser observado no sistema de distribuição. No caso de medições realizadas utilizando-se TPs com conexão do tipo V ou delta aberto, os limites permitidos para o indicador DTT₃₉₅% deverão corresponder a 50% dos respectivos valores indicados na Tabela 8. Os acessantes da Rede Básica deverão seguir o determinado nos Procedimentos de Rede ou em regulamentação específica.

3.5 Considerações Finais

Neste capítulo, apresentou-se os conceitos referentes a QEE, normas técnicas e padrões Brasileiros, Europeus e Americanos os quais são utilizados como padrões de avaliações da QEE no Brasil e no mundo. Bem como as definições sobre os distúrbios harmônicos provocados pela presença de cargas não lineares, o produto energia elétrica, o processo de medição dos harmônicos e os níveis de referência aceitos nos sistemas elétricos. Aborda-se a *Qualidade do Produto*, este define as terminologias, caracteriza os fenômenos e estabelece os indicadores e limites ou valores de referências relativos à conformidade de tensão em regime permanente e às perturbações na forma de onda de tensão.

CAPITULO IV

O PROCESSO DE DESCOBERTA DE CONHECIMENTO EM BASE DE DADOS KDD

4.1 Considerações Iniciais

No cenário mundial, onde os mercados possuem um alto grau de competitividade e concorrência entre as empresas pela busca de conhecimento de novos mercados e produtos, normalmente os dados sobre os possíveis clientes e novos mercados são realizados através de alguma técnica ou sistema de banco de dados. De forma geral as bases de dados são ferramentas indispensáveis na extração de conhecimento.

De acordo com (JAIWEI e KAMBER, 2006), Mineração de Dados é um campo multidisciplinar que inclui áreas como tecnologias de banco de dados, aprendizagem de máquina, estatísticas, reconhecimento de padrões, recuperação de informação, redes neurais, sistemas baseados em conhecimento, inteligência artificial, computação de alto desempenho, e visualização de dados.

KDD é o processo automatizado que se baseia em métodos de diversas áreas, como reconhecimento de padrões, estatísticas, aprendizado de máquina, redes neurais, etc. a encontrar padrões de dados na etapa de mineração de dados. A frase "*Knowledge Discovery in Databases*" foi criada na primeira oficina KDD em 1989 por Piatetsky-Shapiro, que enfatizou que o conhecimento é o produto final de uma descoberta data driven (FRAWLEY, PIATETSKY-SHAPIRO *et al.*, 1991).

Em um nível mais amplo, o que é mineração de dados? É o processo pelo qual a informação buscada é ainda desconhecida, é extraída a partir de grandes volumes de dados. Esta informação deve estar em forma que pode ser entendida, posta em prática, e usado para melhorar os processos de decisão. Obviamente, com essa definição, mineração de dados abrange um amplo conjunto de tecnologias, incluindo armazenamento de dados, gerenciamento de banco de dados, algoritmos de análise de dados e visualização. O ponto crucial do recurso para esta nova tecnologia está nos algoritmos de análise de dados, uma vez que proporcionam mecanismos automatizados para lidar com dados e extrair informações úteis (APTÉ, 1997).

Para (LAKSHMI e RAGHUNANDHAN, 2011) a mineração de dados é uma extração não trivial, implícita de conhecimento e de recurso a partir de grandes conjuntos de dados, é

uma tecnologia em evolução, que é um resultado direto do aumento do uso de bases de dados nos sistemas computacionais, a fim de armazenar e recuperar informações de forma eficaz. Ele também é conhecido como *Knowledge Discovery in Databases* (KDD) e permite a exploração, análise e visualização de dados de grandes bancos de dados em um alto nível de abstração. O trabalho de extração de dados é compreendido por meio de um método chamado de modelagem com objetivo fazer previsões.

Buscando uma forma de mitigar os problemas no contexto de qualidade de energia, foi proposta uma solução com objetivo de utilizar um processo orientado por computador ou processo de descoberta de conhecimento em bases de dados utilizando a técnica bayesianas para analisar os dados oriundos dos campi Básico 1 e 2, Profissional e Saúde, objetivando medir o conteúdo harmônico presente na rede elétrica.

Neste capítulo serão desenvolvidas e aplicadas as características fundamentais do processo de descoberta de conhecimento em base de dados, demonstrando o seu comportamento no transcorrer de cada uma das fases: seleção, pré-processamento, transformação, mineração de dados e interpretação; empregando maior ênfase a etapa de mineração de dados e no algoritmo computacional utilizado que é a técnica empregada (classificador bayesiano *Naive Bayes*) como alvo no processo de aprendizagem.

4.2 Processo do KDD

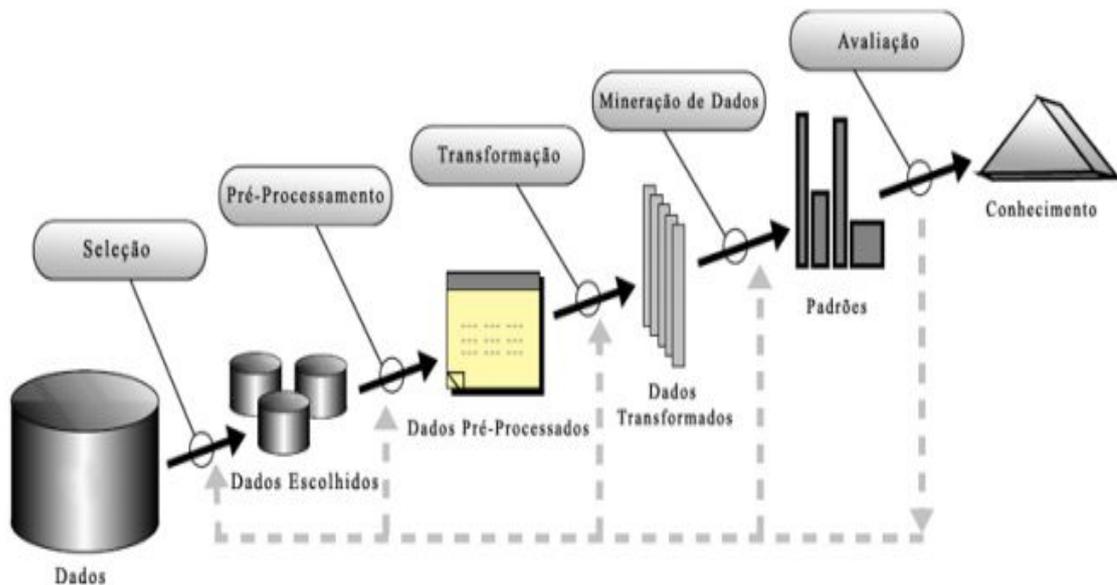
O processo de descoberta de conhecimento é composto por um conjunto de fases que são realizadas para identificação de padrões compreensíveis, validos, novos e potencialmente úteis a partir de base de dados, o processo é iterativo entre as fases e pode ser recursivo. As fases a serem executadas são seleção de dados, pré-processamento, transformação, mineração e análise ou avaliação de resultados.

O processo de descoberta do conhecimento em base de dados é avaliado pela literatura, como sendo uma análise inteligente dos dados, pois buscam e extraem de bases de dados informações que normalmente são desconhecidas e potencialmente úteis.

De acordo com (FAYYAD, 1996), o termo KDD ou Descoberta de Conhecimento em Base de Dados refere-se ao amplo processo de encontrar o conhecimento em dados, e enfatiza a aplicação de "alto nível" de métodos de mineração de dados. Em sua natureza, o KDD engloba as áreas de aprendizagem de máquina, reconhecimento de padrões, bancos de dados, estatísticas, inteligência artificial e visualização de dados.

O processo de KDD é aplicado em fases, no decorrer de sua execução, cada fase deve ser bem aplicada para garantir que os dados sejam bem utilizados e que possam possuir representatividade para as demais fases. Ou seja, quanto melhor forem os dados processados em cada fase, melhores serão os resultados alcançados pelo processo, como demonstrado na Figura 3.

Figura 3 - Processo de KDD.



Fonte: O Processo KDD FAYYAD *et al.*(1996).

4.3 As Fases do Processo de KDD

4.3.1 A Fase de Compreensão do Domínio

É necessário primeiro, desenvolver uma compreensão do domínio do aplicativo e dos antecedentes relevantes. O conhecimento e identificação do objetivo do processo KDD do ponto de vista do cliente (FAYYAD, 1996) prepara a cena para entender o que deve ser feito com as muitas decisões (sobre transformação, algoritmos, representação, etc.). As pessoas que são responsáveis por um projeto KDD precisam entender e definir os objetivos dos usuários finais e do ambiente em que ocorrerá o processo de descoberta de conhecimento (incluindo conhecimento prévio relevante). À medida que o processo KDD prossegue, pode haver até uma revisão e sintonização desta etapa. Tendo entendido os objetivos do KDD, o pré-processamento

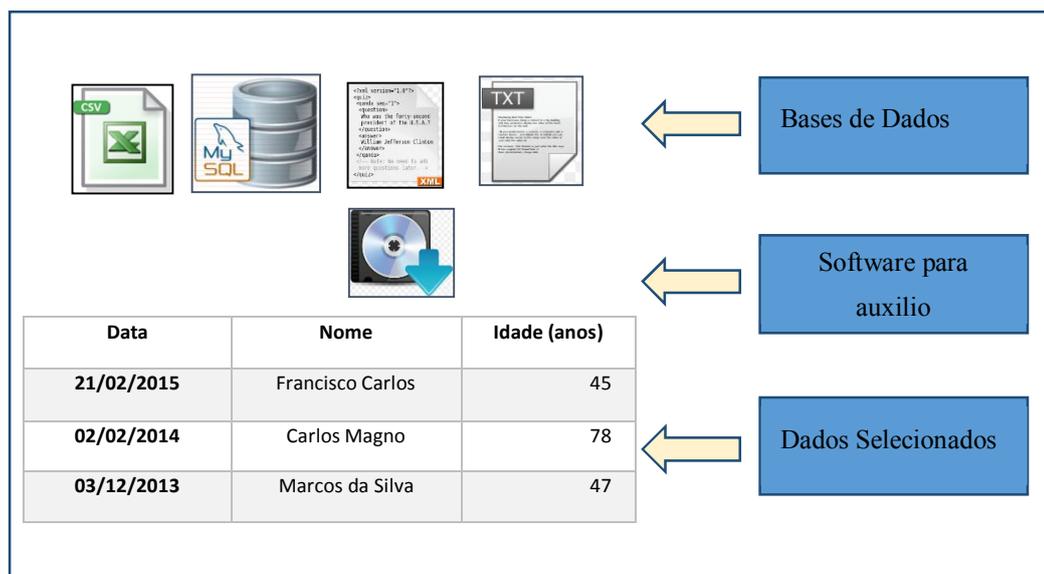
dos dados é iniciado, conforme definido nas próximas três etapas (MAIMON e ROKACH, 2005).

4.3.2 A Fase de Seleção de Dados

A fase de seleção de dados é onde inicia-se o processo de descoberta de descobrimento de informações em bases de dados, esta fase possui grande importância para todos o transcorrer do processo, normalmente esta fase é a que possui o maior impacto no significado dos dados sobre a qualidade do resultado final, uma vez que nela é escolhido o conjunto de dados que contém todas as possíveis características e variáveis que fazem parte das análises. Normalmente a escolha dos dados fica por conta de um especialista de domínio, ou seja, uma pessoa que entenda profundamente do assunto em questão.

O processo de seleção é tido como bastante complexo, pois os dados normalmente estão armazenados em fontes diferentes como planilhas eletrônicas, arquivos de textos, sistemas legados, entre outros. É comum nesta fase ocorrer a criação e o desenvolvimento de softwares para auxiliar no processo de seleção, pois as ferramentas de mineração de dados disponíveis não cobrem todos os casos de formatação dos dados importados de repositórios diferentes, como demonstra a Figura 4.

Figura 4 - Processo de seleção de dados.



Fonte: Autor, (2019).

4.3.3 A Fase de Pré-Processamento

O pré-processamento é a fase onde os dados são submetidos para a limpeza dos dados, eliminação de dados redundantes, eliminação das inconsistências, recuperação de dados incompletos. As execuções destas subfases são cruciais para determinação a qualidade e a eficiência dos algoritmos de mineração de dados das fases seguintes. Assim como as demais fases, é necessário o auxílio de um especialista de domínio, pois na maioria dos casos somente um especialista é capaz de analisar os dados e definir se os dados pré-processados são dados utilizáveis. Ainda nesta fase, são usados métodos de redução e ou transformação de dados, com o objetivo de diminuir a quantidade de variáveis utilizadas no processo, visando com isto melhorar o desempenho e as análises a serem realizadas pelos algoritmos.

A discretização de dados é um método de pré-processamento de finalidade geral que reduz o número de valores distintos para uma determinada variável contínua, dividindo a sua gama em um conjunto finito de intervalos disjuntos, e então relaciona esses intervalos com rótulos significativos (LINOFF e BERRY, 2011). Posteriormente, os dados são analisados ou relatados neste nível superior de representação do conhecimento, em vez de os valores individuais sutis e, portanto, leva a uma representação simplificada na exploração de dados e processos de mineração. Um processo de discretização flui em quatro passos (KOTSIANTIS e KANELLOPOULOS, 2006) como representado a seguir:

1. Classifica os valores contínuos do atributo a serem discretizados;
2. Avalia um ponto de corte para intervalos de divisão ou adjacentes à fusão;
3. De acordo com alguns critérios, executa-se a divisão ou fusão de intervalos de valores contínuos;
4. Por último estabelece a parada em algum ponto com base em critérios de parada.

O objetivo da discretização é encontrar um conjunto de pontos de corte para particionar em um pequeno número de intervalos. A motivação para a discretização de dados baseia-se na necessidade de obter maiores taxas de precisão a fim de manipular dados com atributos de alta cardinalidade e evitar problemas de algoritmos que não tratam variáveis contínuas diretamente. Geralmente, os métodos de discretização podem ser classificados como: supervisionado ou não supervisionado, direto ou incremental, global ou local, estático ou dinâmico, *topdown* ou *bottom-up* (KOTSIANTIS e KANELLOPOULOS, 2006).

Neste trabalho utiliza-se a discretização não supervisionada de igual frequência, por se mostrar o método que gerou melhores resultados para os experimentos. O algoritmo de igual

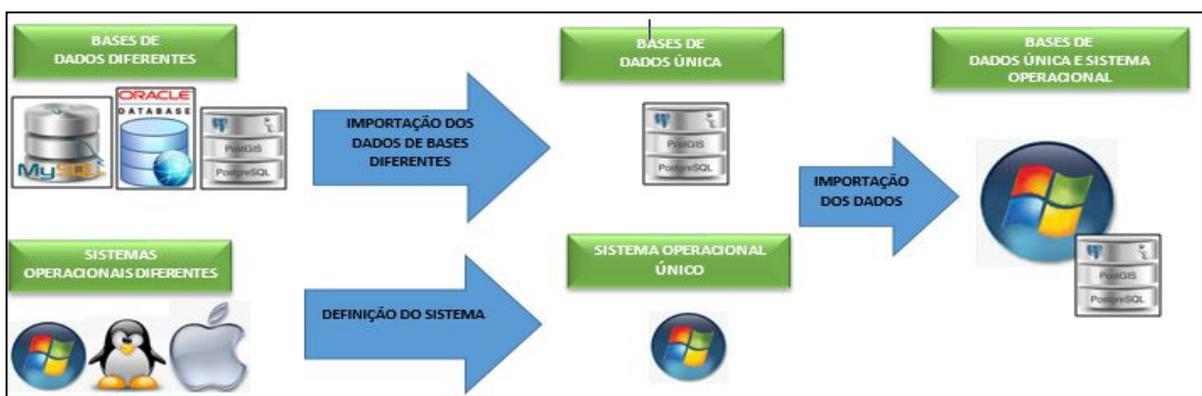
frequência determina os valores mínimos e máximos do atributo de dados discretos, classifica todos os valores em ordem crescente, e divide os valores contínuos classificados em intervalos k tal que cada intervalo contém aproximadamente n/k instâncias de dados com valores adjacentes. Para igual frequência, muitas ocorrências de um valor contínuo podem causar as ocorrências a serem atribuídos em diferentes faixas. Este algoritmo tenta superar as limitações da discretização intervalo de largura igual dividindo o domínio em intervalos com a mesma distribuição de pontos de dados. Os exemplos de dados com valores idênticos devem ser colocados no mesmo intervalo de tempo, pelo que nem sempre é possível gerar exatamente k intervalos iguais de frequência. Este método também é chamado como discretização proporcional de k -intervalos (DASH, PARAMGURU *et al.*, 2011).

4.3.4 A Fase de Transformação

A quarta fase é a redução e projeção de dados, encontrar recursos úteis para representar os dados de acordo com o objetivo da tarefa. Com métodos de redução da dimensionalidade ou de transformação, o número efetivo de variáveis em consideração pode ser reduzido, ou podem ser encontradas representações invariantes para os dados (AL-GARADI, KHAN *et al.*, 2016).

A transformação dos dados é a fase do KDD que ocorre antes da fase de mineração de dados. Após os dados terem passado pelos processos de limpeza e pré-processamento, os mesmos necessitam serem armazenados e formatados adequadamente para que os algoritmos de inteligência computacional possam ser aplicados.

Figura 5 - Processo de Armazenamento em Banco de Dados e Sistema Operacional.



Fonte: Autor, (2019).

Na fase transformação é comum que os dados se encontrem em computadores diferentes, sistemas operacionais diferentes e em sistemas de gerenciamento de bases de dados

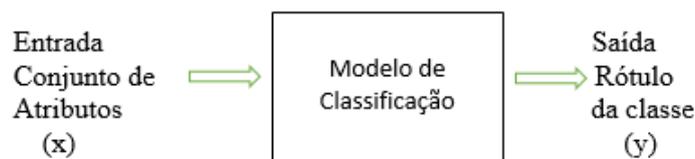
(SGBD) diferentes que causam grandes problemas no momento em que estes devem ser unificados em uma única base de dados, nesta fase os dados devem necessariamente passar pelo processo de unificação de bases de dados, ou seja, todos os dados serão agrupados em um repositório único, com o objetivo de melhorar o acesso e a velocidade de acesso dos algoritmos de aprendizagem utilizados no processo, como apresentado na Figura 5. Além disto, nesta fase, também é possível obter os dados incompletos ou faltantes através da transformação ou combinação de outros dados existentes que podem gerar os dados conhecidos como “dados derivados”.

4.3.5 A Fase de Objetivos do Processo KDD

A quinta fase corresponde aos objetivos do processo KDD (passo 1) para um método particular de mineração de dados. Por exemplo, sumarização, classificação, regressão, agrupamento, etc., (FAYYAD, 1996). A classificação é uma forma de análise de dados que extrai modelos que descrevem classes importantes de dados. Tais modelos, chamados classificadores, predizem rótulos de classe categóricos (discretos, não ordenados), (JAIWEI e KAMBER, 2006) .

A classificação é a tarefa de organizar objetos em uma entre diversas categorias pré-definidas, é um problema universal que engloba muitas aplicações diferentes.

Figura 6 - Classificação como a tarefa de mapear um conjunto de atributos x no seu rótulo de classe y .



Fonte: Tan, Steinbach *et al.*, (2009).

Definição de Classificação: Classificação é a tarefa de aprender uma função alvo f que mapeie cada conjunto de atributos x para um dos seus rótulos de classes y pré-determinados. A função alvo também é conhecida informalmente como modelo de classificação. Um modelo de classificação é útil para os seguintes propósitos, modelagem descritiva e modelagem preditiva (TAN, STEINBACH *et al.*, 2009).

4.3.6 A Fase de Análise Exploratória e Seleção de Modelo e Hipótese

A Sexta fase é a análise exploratória e seleção de modelo e hipótese: escolhendo o(s) algoritmo(s) de data mining e os métodos de seleção a serem usados para pesquisar padrões de dados. Essa fase é essencial pois o algoritmo terá total influência sobre os resultados obtidos na fase de mineração de dados. Este processo inclui decidir quais modelos e parâmetros podem ser apropriados (por exemplo, os modelos de dados categóricos são diferentes dos modelos de vetores sobre os reais) e combinando um método particular de mineração de dados com os critérios gerais do processo KDD (por exemplo, o usuário final pode estar mais interessado em entender o modelo do que suas capacidades preditivas) (FAYYAD, 1996).

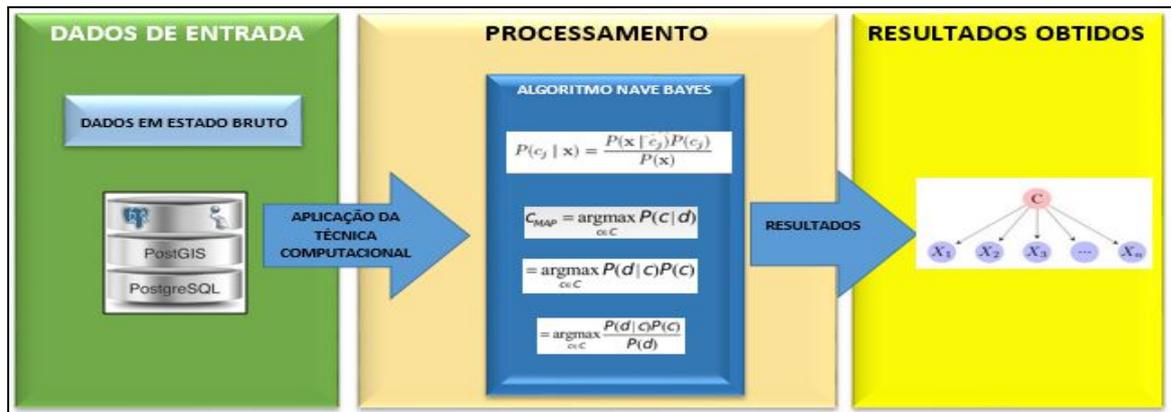
Para (MAIMON e ROKACH, 2005), nesta fase se escolher o algoritmo Mineração de Dados. Tendo a estratégia, agora decidimos as táticas. Esta etapa inclui a seleção do método específico a ser usado para pesquisar padrões (incluindo indutores múltiplos). Por exemplo, ao considerar a precisão versus a compreensão, o primeiro é melhor com as redes neurais, enquanto o segundo é melhor com árvores de decisão. Para cada estratégia de meta-aprendizagem, existem várias possibilidades de como isso pode ser realizado. O Meta-aprendizados concentra-se em explicar o que faz com que um algoritmo de Data Mining seja bem-sucedida ou não em um problema específico. Assim, essa abordagem tenta entender as condições sob as quais um algoritmo de Mineração de dados é mais apropriado. Cada algoritmo possui parâmetros e táticas de aprendizagem como a validação cruzada em dez vezes ou outra divisão para treinar e testar.

4.3.7 A FASE DE MINERAÇÃO DE DADOS

De acordo com (FAYYAD, 1996), a sétima fase é a mineração de dados que busca por padrões de interesse em uma determinada forma de representação ou um conjunto dessas representações, incluindo regras de classificação ou árvores, regressão e agrupamento. O usuário pode ajudar significativamente o método de mineração de dados executando corretamente as etapas anteriores.

Entre todas as etapas do processo de KDD a mineração de dados é a mais afetada pelas fases anteriores, é conhecida como a fase mais importante, no entanto, ela depende muito das fases anteriores para que seja bem-sucedida. Caso as fases anteriores não sejam muito bem desenvolvidas, esta fase poderá apresentar problemas na aplicação dos algoritmos causando assim vários problemas, a figura 7 apresenta um esquema da fase de mineração de dados.

Figura 7 - Processo de mineração de dados.



Fonte: Autor, (2019).

Quanto melhores forem as fases anteriores, melhores serão os resultados obtidos nesta fase, que é reconhecida pela literatura como a mais importante das fases do processo de KDD.

4.3.8 A Fase de Interpretação e avaliação

Segundo (FAYYAD, 1996), o processo KDD global figura 3 inclui a avaliação e a possível interpretação dos padrões minerados para determinar quais padrões podem ser considerados novos conhecimentos. É a fase onde os resultados alcançados no decorrer do processo são analisados, interpretados e verificados por um especialista de domínio, nesta fase é determinado se o conhecimento adquirido pode ser utilizado e se possui alguma validade. Identificado os padrões pelo sistema, estes são interpretados em conhecimentos, os quais darão suporte a tomada de decisões humanas.

Para (SUMATHI e SIVANANDAM, 2006) a mineração de dados necessita das pessoas certas, independentemente do que muitos produtores de ferramentas de mineração de dados afirmam, a mineração de dados não é (ainda) uma operação "automática" com pouca ou nenhuma intervenção humana. Pelo contrário, o analista humano desempenha um papel importante, principalmente nas áreas de seleção de dados e interpretação de dados / conhecimento.

4.3.9 A Fase do Conhecimento

De acordo com (FAYYAD, 1996), a nona fase atua no conhecimento descoberto, usando o conhecimento diretamente, incorporando o conhecimento em outro sistema para ações

futuras, ou simplesmente documentando-o e informando as partes interessadas. Esse processo também inclui verificar e resolver possíveis conflitos com o conhecimento acreditado anteriormente (ou extraído).

Figura 8 - Fase de Interpretação do KDD.



Fonte: Autor, (2019).

Usando o conhecimento descoberto, agora é possível incorporar o conhecimento em outro sistema para novas ações. O conhecimento torna-se ativo no sentido de que pode-se fazer mudanças no sistema e medir os efeitos. Há muitos desafios nesta etapa, como perder as "condições de laboratório" sob as quais se opera. Por exemplo, o conhecimento foi descoberto a partir de um determinado instantâneo estático (geralmente amostra) dos dados, mas agora os dados se tornam dinâmicos. As estruturas de dados podem mudar (certos atributos ficam indisponíveis) e o domínio de dados pode ser modificado (como, um atributo pode ter um valor que não foi assumido anteriormente).

4.4 Considerações Finais

Neste capítulo, abordou-se o processo de extração de conhecimento em bases de dados (KDD), o qual tem por objetivo básico, adquirir conhecimento através de bases de dados, descobrir padrões, regras e relações contidas na base de dados, fornecer aos usuários, métodos e técnicas que auxiliem nos processos de avaliações e tomadas de decisões.

CAPITULO V

NAIVE BAYES

5.1 Classificador

Um mapeamento de instâncias não marcadas para classes (discretas). Os classificadores têm um formulário (por exemplo, árvore de decisão) mais um procedimento de interpretação (incluindo como lidar com incógnitas, etc.). Alguns classificadores também fornecem estimativas de probabilidade (pontuações), que podem ser ajustadas para produzir uma decisão de classe discreta levando em consideração uma função de utilidade (R. e F.).

Para (MAIMON e ROKACH, 2005), o classificador gerado pelo indutor pode ser usado para classificar uma tupla não vista, atribuindo-a explicitamente a uma determinada classe (classificador nítido) ou fornecendo um vetor de probabilidades que representa a probabilidade condicional da instância dada para pertencer a cada classe (classificador probabilístico).

Supondo-se que para um universo de objetos é definido que:

Um conjunto de variáveis X_1, X_2, \dots, X_n , chamadas atributos que descrevem estes objetos e uma variável C chamada classe cujos valores c_1, \dots, c_m representam categorias ou classes que possam ser atribuídas a estes objetos.

Um classificador é uma função que atribui uma classe a objetos descritos por um conjunto de atributos, como mostra a figura 9 (CASTILLO, 2012).

Figura 9 - Esquema de um classificador



Fonte: 9 (CASTILLO, 2012).

Matematicamente:

Um classificador é uma função $f: X \rightarrow C$, que atribui uma classe $c \in C$ a objetos (exemplos) $x = (x_1, x_2, \dots, x_n) \in X$.

5.2 Aprendizagem Supervisada

A abordagem mais popular para a aprendizagem é através do exemplo. Dado um conjunto de objetos, cada um marcado com uma classe (categoria), o sistema de aprendizagem constrói um mapeamento entre objetos e classes que podem então ser utilizados para classificação de novos objetos (não marcados).

Como a rotulagem (categorização) da (formação) do conjunto inicial de objetos é feita por um agente externo ao sistema (professor), essa configuração é chamada aprendizagem supervisionada (MARKOV e LAROSE, 2007).

Na aprendizagem supervisionada, um conjunto de padrões de treinamento é usado onde o rótulo de classe para cada padrão de dado é usado para classificação. O conjunto de treinamento completo pode não ser usado porque o tempo de processamento pode ser muito longo, mas uma abstração do conjunto de treinamento pode ser usada.

No aprendizado supervisionado, a eficácia da classificação depende dos dados de treinamento. Um dado de treinamento maior pode levar a uma melhor classificação, mas pode exigir mais tempo para o processo de classificação (MURTY e DEVI, 2011).

A palavra “supervisada” (do inglês supervised) refere-se a que os objetos usados para a tarefa de aprendizagem foram classificados previamente por um supervisor.

5.3 Definições Básicas de Probabilidades

A Probabilidade é um campo da matemática que estuda e analisa a ocorrência de fenômenos aleatórios. Fenômenos aleatórios são experimentos repetidos sob as mesmas condições que produzem resultados que não se pode prever com certeza (MORGADO et al., 2001). Outros conceitos importantes dentro da probabilidade são definidos a seguir.

- Espaço amostral - O espaço amostral de um experimento aleatório é o conjunto de todos os resultados possíveis desse experimento. Pode-se denotar tal conjunto pela letra grega *ômega maiúscula*, Ω . Quando o espaço amostral é finito ou infinito enumerável, é chamado espaço amostral discreto. Caso contrário, isto é, quando Ω é não-enumerável, chama-se de espaço amostral contínuo.
- Eventos aleatórios - Os subconjuntos de Ω são chamados *eventos aleatórios*; já os elementos de Ω são chamados eventos elementares. A classe dos eventos aleatórios de

um espaço amostral, que denotaremos por $F(\Omega)$, é o conjunto de todos os eventos (isto é, de todos os subconjuntos) do espaço amostral.

A título de ilustração, consideremos um espaço amostral com três elementos: $\Omega = \{\omega_1, \omega_2, \omega_3\}$. A classe dos eventos aleatórios é

$$F(\Omega) = \{\emptyset, \{\omega_1\}, \{\omega_2\}, \{\omega_3\}, \{\omega_1, \omega_2\}, \{\omega_1, \omega_3\}, \{\omega_2, \omega_3\}, \{\omega_1, \omega_2, \omega_3\}\}$$

Os eventos, sendo conjuntos, serão representados por letras maiúsculas do nosso alfabeto, enquanto os elementos de um evento serão representados por letras minúsculas.

- Operações com eventos aleatórios:

$$\text{Interseção } x \in A \cap B \Leftrightarrow x \in A \text{ e } x \in B$$

$$\text{União } x \in A \cup B \Leftrightarrow x \in A \text{ ou } x \in B$$

$$\text{Exclusão } A \cap B = \emptyset$$

$$\text{Complementar } x \in \neg A \Leftrightarrow x \notin A$$

$$\text{Diferença } x \in A - B \Leftrightarrow x \in A \text{ e } x \notin B \quad A - B = A \cap \neg B$$

$$\text{Partição } A_1, A_2, \dots, A_n : A_i \cap A_j = \emptyset \text{ e } \bigcup_i A_i = \Omega$$

- Partição de um espaço amostral

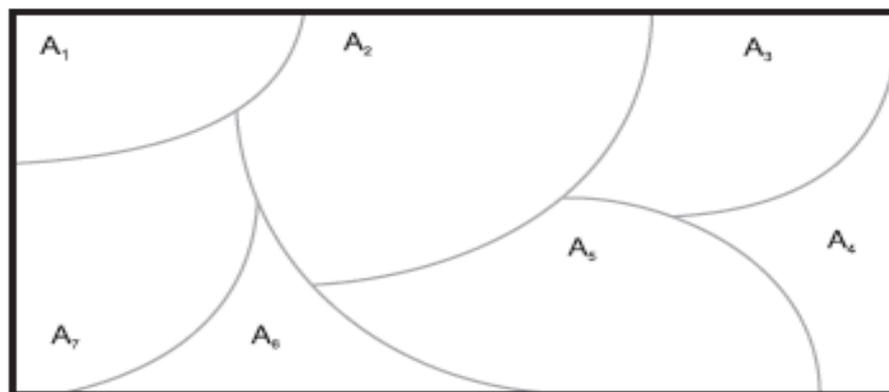
Uma coleção de eventos A_1, A_2, \dots, A_n forma uma partição do espaço amostral Ω se:

I. Os eventos A_i são disjuntos dois a dois, isto é, se $A_i \cap A_j = \emptyset \forall i \neq j$;

II. A união dos eventos A_i é o espaço amostral Ω , isto é, $\bigcup_{i=1}^n A_i = \Omega$.

Na figura 10 ilustra-se esse conceito.

Figura 10 - Partição do espaço amostral Ω .



Fonte: Farias (2010)

A probabilidade de um evento A é denotada por $P(A)$ e toda função de probabilidade P deve satisfazer os três axiomas:

- (1) A probabilidade de todo o espaço amostral é 1, i.e., $P(\Omega) = 1$;
- (2) A probabilidade de qualquer evento elementar é não negativa, ou seja, $\forall A \in \Omega$, $0 \leq P(A) \leq 1$;
- (3) Se n eventos A_1, A_2, \dots, A_n são mutuamente exclusivos, ou seja, eles não podem ocorrer simultaneamente, então a probabilidade de que pelo menos um destes n eventos ocorra é dada pela soma das probabilidades de cada um deles, como mostra a Eq. (4).

$$P(A_1 \cup A_2 \cup \dots \cup A_n) = \sum_{i=1, n} P(A_i) \quad (4)$$

Os axiomas (1) e (2) podem ser combinados na expressão em Eq. (5).

$$\forall A \in \Omega, 0 \leq P(A) \leq 1 \quad (5)$$

O complemento de A , ou negação do evento A ($\neg A$) contém a coleção de todos os eventos elementares no espaço amostral Ω , exceto o evento A . Considerando que A e $\neg A$ são mutuamente exclusivos e $A \cup \neg A = \Omega$, pelos axiomas (1) e (3) tem-se que:

$$P(A \cup \neg A) = P(A) + P(\neg A) = P(\Omega) = 1, \text{ i.e.,} \quad (6)$$

$$P(\neg A) = 1 - P(A)$$

5.4 Naive Bayes

Minsky (1961) mostra que, se as variáveis preditoras aleatórias, como variáveis de classe são binárias, a superfície de decisão que é derivada de um modelo Naive Bayes é um hiperplano (MINSKY, 1961).

Apesar de ter uma longa tradição na comunidade de reconhecimento de padrões (Duda e Hart, 1973), o classificador *Naive Bayes* aparece pela primeira vez na literatura de aprendizagem de máquina em finais dos anos oitenta (Cestnik et al. 1987), a fim de comparar a sua capacidade preditiva com métodos mais sofisticados (LARRAÑAGA, INZA *et al.*, 1997).

Aos poucos, os investigadores desta comunidade de aprendizagem de máquina já perceberam o seu potencial e robustez em problemas de classificação supervisionada.

Redes de Bayesianas são ferramentas poderosas para decisão e raciocínio sob incerteza. Uma forma muito simples de rede Bayes é chamado de classificadores Naive Bayes dedicados à tarefa de classificação (DUDA, HART *et al.*, 1973). Essas redes são compostas por grafos acíclicos direcionados com uma estrutura única, tem um nó raiz (chamado pai), o que representa o nó de classe, e vários filhos, o que corresponde a atribuir nós. Naive Bayes trabalha sob a forte suposição ‘ingênua’ de independência entre nós filhos no contexto de seu pai. Entretanto, essa suposição raramente é verdadeira em problemas reais. Assim, na presença de um conjunto de treinamento só devemos calcular probabilidades condicionais de um modo uma vez que a estrutura é única.

O Classificador Naive Bayes (DUDA, HART *et al.*, 1973) baseia-se na suposição de independência condicional entre os atributos dada a classe. Entretanto, essa suposição raramente é verdadeira em problemas reais, embora o algoritmo Naive Bayes tenha se mostrado competitivo quando comparado a outros algoritmos mais complexos.

O algoritmo Naive Bayes apresenta uma maneira de combinar as probabilidades prévias e probabilidades condicionais em uma única fórmula, que pode ser usada para calcular a probabilidade de cada uma das possíveis classificações por sua vez. Tendo feito isso, escolhemos a classificação com o maior valor (BRAMER, 2016).

Segundo (IYER, SHAFRAN *et al.*, 2013), estas suposições geram algoritmos bastante eficientes quer para a classificação quer para a aprendizagem. Por exemplo, seja C uma variável aleatória que expressa a classe de uma instância e X o vetor de variáveis aleatórias que expressam os valores dos atributos observáveis. Seja c a representação de uma classe em particular, e x a representação de um valor observável em particular. Dado um caso de teste x a classificar, basta usar a regra de Bayes para calcular a probabilidade de x pertencer a cada classe c , o que permite encontrar a classe com probabilidade de maior ordem e tomar uma decisão quanto à classificação desse caso de teste e justamente por se escolher essa classe com maior probabilidade.

Para (MITCHELL, 1997), para classificar uma nova instância, o algoritmo determina a classe mais provável, dados os atributos (a_1, a_2, \dots, a_M) que descrevem a instância. A equação 7 mostra o cálculo da classe de maior probabilidade para o classificador *Naive Bayes*, onde V_{NB} simboliza a resposta do classificador, $P(v_j)$ corresponde à frequência estimada de instâncias de treinamento que pertencem a cada classe v_j , e $P(a_i|v_j)$ é a frequência estimada dos valores do atributo a_i restrito aos exemplos de treinamento da classe v_j .

$$v_{NB} = \arg \max_{v_j \in V} P(v_j) \prod_i P(a_i | v_j) \quad (7)$$

Para (KJÆRULFF e MADSEN, 2013), um modelo Naive Bayes é caracterizado por uma estrutura em que todas as variáveis, são filhos (variáveis de atributo) de uma única variável de *classe* e as variáveis de atributo são condicionalmente independentes, dada a variável de classe. Isso torna a construção do modelo muito fácil e a inferência é muito eficiente. O modelo é popular para resolver problemas de classificação.

Use:

- Usa-se o modelo Bayes ingênuo se você for confrontado com um problema de classificação ou (equivalente) um problema de diagnóstico com uma única variável que represente um conjunto de diagnósticos mutuamente exclusivos com sintomas condicionalmente independentes desde o diagnóstico.

Não:

- Se usa o modelo Bayes ingênuo se dois ou mais problemas (por exemplo, classes em um problema de classificação ou diagnósticos em um problema de diagnóstico) podem coexistir (isto é, não são mutuamente exclusivos).
- Se usa o modelo Naive Bayes se as variáveis de atributos (por exemplo, recursos em um problema de classificação ou sintomas em um problema de diagnóstico) são (fortemente) condicionalmente dependentes, dada a classe ou variável de diagnóstico.
- Se usa o modelo Naive Bayes se o modelo de probabilidade for suposto ser exato, já que a distribuição de probabilidade conjunta do modelo Bayes nativo tende às vezes não ser bem calibrada.

A técnica de modelagem Naive Bayes é implementada da seguinte forma:

1. Seja C a variável de classe com um estado para cada classe possível.
2. Seja $J = \{I_1, \dots, I_n\}$ o conjunto de variáveis recurso.
3. Deixar C não ter pais e deixá-lo ser o único pai de cada característica variável I_i .

Desta maneira, a variável C torna-se intermediário de uma ligação em série com cada par de variáveis de recursos.

4. Definir a distribuição de probabilidade anterior $P(C)$ que codifica tal a frequência relativa de cada classe.

5. Para cada $I_i \in J$ define-se a distribuição de probabilidade condicional $P(I_i | C)$ tal que ele codifica a frequência relativa de cada estado do estado de cada recurso dada a variável de classe. Se os dados disponíveis em seguida, apresentam uma vantagem para estimar a

distribuição condicional de probabilidade prévia $P(C)$ e $P(I_1 | C), \dots, P(I_n | C)$ a partir dos dados (KJAERULFF e MADSEN, 2012).

O Teorema de Bayes calcula a probabilidade de diferentes hipóteses à medida que novas evidências são observadas.

$$P(A \cap B) = P(A | B) * P(B) \quad (8)$$

Por outro lado, a probabilidade de que ocorra simultaneamente um evento A e um evento B também pode ser dada por:

$$P(B \cap A) = P(B | A) * P(A) \quad (9)$$

Combinando (8) e (9), tem-se:

$$P(A | B) * P(B) = P(B | A) * P(A) \quad (10)$$

Rearranjando, chega-se ao teorema de Bayes

$$P(A|B) = \frac{P(B|A)*P(A)}{P(B)} \quad (11)$$

onde:

A representa qualquer *hipótese* cuja probabilidade pode ser afetada por dados (chamada evidência abaixo). Muitas vezes, há hipóteses concorrentes, a partir do qual se escolhe a mais provável.

A evidência B , corresponde a novos dados que não foram utilizados no cálculo da probabilidade anterior.

$P(A)$, a probabilidade a priori, é a probabilidade de A antes de B ser observada. Isso indica sua estimativa anterior da probabilidade de que uma hipótese seja verdadeira, antes de obter a evidência atual.

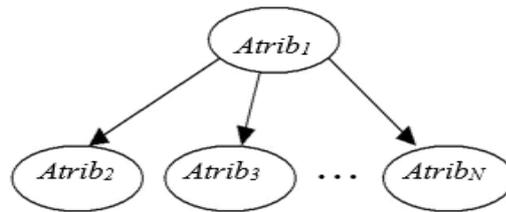
$P(A|B)$ a probabilidade a posteriori, é a probabilidade de A dado B , ou seja, depois de B ser observada. Isto indica o que se quer saber: a probabilidade de uma hipótese dada a evidência observada.

$P(B|A)$ é a probabilidade de observar B dado A . Indica a compatibilidade dos elementos com a hipótese dada.

$P(B)$ é às vezes chamado de probabilidade marginal ou "evidência modelo". Este fator é a mesma para todas as hipóteses possíveis que estão sendo considerados. Isto significa que este fator não entra em determinar as probabilidades relativas de diferentes hipóteses.

No *Naive Bayes*, todos os atributos da base de dados são considerados independentes entre si, mas sendo dependentes de um determinado *nó pai* (um dos atributos da base de dados escolhidos como principal, pelo qual todos os outros atributos restantes apresentam uma certa dependência), como pode ser visto na figura 11 (SANTANA, 2005).

Figura 11 - Modelo de um Naive Bayes.



Fonte: Santana (2005)

Uma vez considerada a independência entre os eventos, ou seja, não existe nenhuma relação causal entre eles, o cálculo de probabilidade da inferência aplicada na rede é realizado segundo a aplicação da equação (12).

$$P(x_i | c_1, c_2, c_3, \dots, c_n) = P(x_i) \prod_{k=1}^n P(c_k | x_i) \quad (12)$$

Na qual tenta-se calcular a probabilidade de um determinado evento (x_i) ocorrer baseado na evidência de outros eventos (c_1, c_2, \dots, c_n) também ocorrerem (SANTANA, 2005).

5.5 Considerações Finais

Neste capítulo fiz-se uma breve introdução ao cálculo das probabilidades e apresentou-se o algoritmo Naive Bayes, um tipo específico de Rede Bayesiana simples, apresentando suas características onde todos os atributos são considerados independentes entre si, mas dependentes de um atributo em particular (pai), por ser um algoritmo que é robusto, simples,

computacionalmente eficiente e cujas características foram propicias para uma execução deste trabalho.

CAPÍTULO VI

APLICAÇÃO DO KDD NA ANÁLISE DA QUALIDADE DE ENERGIA: ESTUDO DE CASO

6.1 Introdução

Atualmente desenvolve-se projeto que tem a finalidade de desenvolver um Sistema de Gestão (Sistema de Gestão das Unidades Consumidoras de Energia Elétrica- SISGEE) no ambiente de pesquisa do Centro de Excelência em Eficiência Energética da Amazônia CEAMAZON, o qual é uma unidade do ITEC - Instituto de Tecnologia da Universidade Federal do Pará, que se propõe a realizar uma campanha de medição nas instalações consumidoras situadas na Cidade Universitária Prof. José da Silveira Neto, com o objetivo de emitir um diagnóstico energético sobre as condições das instalações elétricas, e sobre a qualidade da energia dessas instalações.

O presente estudo foi desenvolvido na Universidade Federal do Pará (UFPA), no Campus Prof. José da Silveira Neto, conhecido como “Campus Universitário do Guamá”, localizado na cidade de Belém/PA, às margens do rio Guamá, é identificado pela concessionária de distribuição de energia elétrica local como Unidade Consumidora (UC) 19.

O Campus Prof. José da Silveira Neto é considerado como a principal unidade consumidora de energia elétrica da Universidade Federal do Pará, a UC-19 engloba 84 (oitenta e quatro) unidades acadêmicas e/ou administrativas ao longo do Campus Universitário do Guamá e tem o seu fornecimento de energia elétrica em 13.8 kV, através do alimentador Guamá GM-06, proveniente da Subestação Guamá da concessionária de energia, Centrais Elétricas do Pará (CELPA). Todas as suas edificações são atendidas pelos circuitos oriundos da cabine de medição geral, localizada no Campus Básico, onde é instalada a proteção primária geral desta unidade consumidora.

6.2 Descrição do Sistema Elétrico do Campus do Guamá

A Prefeitura do Campus Universitário (PCU), realizou Projeto de Modernização da Rede Elétrica da UFPA campus do Guamá, a rede de distribuição aérea protegida compacta é uma

solução tecnológica que possibilita melhorar os níveis de qualidade da energia distribuída e a confiabilidade do sistema. As subestações são predominantemente aéreas, possuindo ainda algumas subestações abrigadas, as quais são alimentadas por ramais subterrâneos a partir da rede aérea.

A rede compacta é um conjunto formado por cabo de aço e cabos cobertos, fixados em estruturas compostas por braços metálicos, espaçadores losangulares ou separadores de fase confeccionados em material polimérico. Todo o conjunto é sustentado pelo cabo de aço aluminizado que, além da função de sustentar o conjunto, compartilha também como neutro do sistema de distribuição, denominado, nesta padronização, de cabo mensageiro/neutro (MATOS, 2016).

Partindo do ponto de entrega e medição de energia, o atendimento interno é feito em rede de distribuição primária através dos quatro alimentadores (AL), identificados como AL-01, AL-02, AL-03 e AL-04, distribuídos entre os setores a seguir relacionados, básico 1 (AL-01), básico 2 (AL-02), setor profissional (AL-03) e o setor saúde (AL-04). O sistema elétrico é constituído de uma rede predominante aérea protegida compacta (rede compacta – Spacer Cable) e caracteriza-se por uma concepção radial de alimentadores.

A cabine de medição geral destes alimentadores possui um sistema de seccionamento e proteção para o alimentador, constituído de uma chave seccionadora sob carga instalada a montante de um disjuntor a vácuo, acionado, remotamente, a partir de uma sala destinada ao sistema de proteção, o qual é constituído por um rele digital multifunção. Na Tabela 11 demonstra-se as especificidades de cada um dos alimentadores (MATOS, 2016).

Tabela 11 - Os Alimentadores da Universidade Federal do Pará

ALIMENTADOR	COMPRIMENTO	LOCALIZAÇÃO	Nº UNIDADES ATENDIDAS	POTENCIA (KVA)
AL – 01	1.682,56 m	CAMPUS BÁSICO 1	30	5.587,50
AL – 02	1269,19 m	CAMPUS BÁSICO 2	13	3.775
AL – 03	2.055 m	CAMPUS PROFISSIONAL	26	5.950
AL – 04	2.959,53 m	CAMPUS SAÚDE	15	3.012,50

Fonte: Matos (2016).

O estudo foi desenvolvido e testado na rede elétrica de distribuição do Campus Universitário do Guamá, com o objetivo de identificação de fontes harmônicas e seus impactos no Ponto de Acoplamento Comum entre a concessionária de energia elétrica e o consumidor.

Os dados foram coletados nesta universidade a qual está dívida em campus conforme descrito na tabela 11 com seus prédios, laboratórios e demais instalações que geram os impactos de distorções harmônicas na rede, um dos grandes problemas analisado.

As coletas de dados foram realizadas nos seguintes pontos, o primeiro medidor foi colocado na entrada principal e os outros quatro conforme descrito na tabela 11, o que totalizou a quantidade de cinco medidores. Nestes pontos, foram instalados medidores de qualidade de energia HIOKI PW3198.

6.2.1 Equipamentos Utilizados nas Medições

No Campus Universitário analisado, instalou-se cinco (5) analisadores de QEE que possuíam as características descritas na tabela 12, os quais foram instalados e monitoraram os pontos descritos anteriormente. Realizou-se a campanha de medição de uma semana conforme estabelecido pela ANEEL (Agência Nacional de Energia Elétrica), PRODIST no módulo 8 no item 9 (INSTRUMENTAÇÃO E METODOLOGIA DE MEDIÇÃO).

Tabela 12 - Características dos analisadores de QEE

Tipo de linha de Medição	Monofásica 2 fios, monofásico 3 fios trifásico 3 fios ou trifásica 4 fios mais um canal de entrada extra (deve ser sincronizado para referenciar canal durante a medição AC / DC)
Faixas de tensão	Medição de tensão: 600 V rms
Faixas de corrente	Medida transitória 6,0000 kV pico
Exatidão básica	500,00 mA para 5,0000 kA AC (depende do sensor de corrente em uso)
Itens de medição	Tensão: $\pm 0,1\%$ da tensão nominal
Registro	Atual: $\pm 0,2\%$ leitura. $\pm 0,1\%$ F.S. + Precisão do sensor de corrente
Interfaces	Potência ativa: $\pm 0,2\%$ leitura. $\pm 0,1\%$ F.S. + Precisão do sensor de corrente
Exibição	1. Transitório sobre tensão: MHz amostragem 2
Fonte de energia	Ciclo 2. Frequência: Calculado como um ciclo, de 40 a 70 Hz
Dimensões e massa	3. Tensão (1/2) RMS: um cálculo ciclo atualizado a cada meio ciclo
Acessórios	Atuais (1/2) RMS: cálculo semicírculo

Fonte: O Autor (2019)

O PW3198 é um analisador de qualidade de energia para monitoramento e gravação de anomalias de fornecimento de energia, permitindo que suas causas sejam identificadas, e também para avaliar os problemas de fornecimento de energia, tais como quedas de tensão,

flicker, harmônicos e outros problemas elétricos. A figura 12 apresenta uma imagem ilustrativa do mesmo.

Figura 12 - Imagem ilustrativa do Analisador de QEE HIOKI PW3198



Fonte: HIOKI (2019)

6.3 Metodologia Utilizada para Análise dos Dados

A metodologia computacional utilizada para auxiliar o diagnóstico da qualidade de energia elétrica sob a ótica das distorções harmônicas, nas análises dos dados coletados pelos analisadores, foi o processo de KDD. O processo ocorre em etapas as quais serão descritas a seguir, aplicando o classificador de Naive Bayes.

6.3.1 Desenvolver uma Compreensão do Domínio

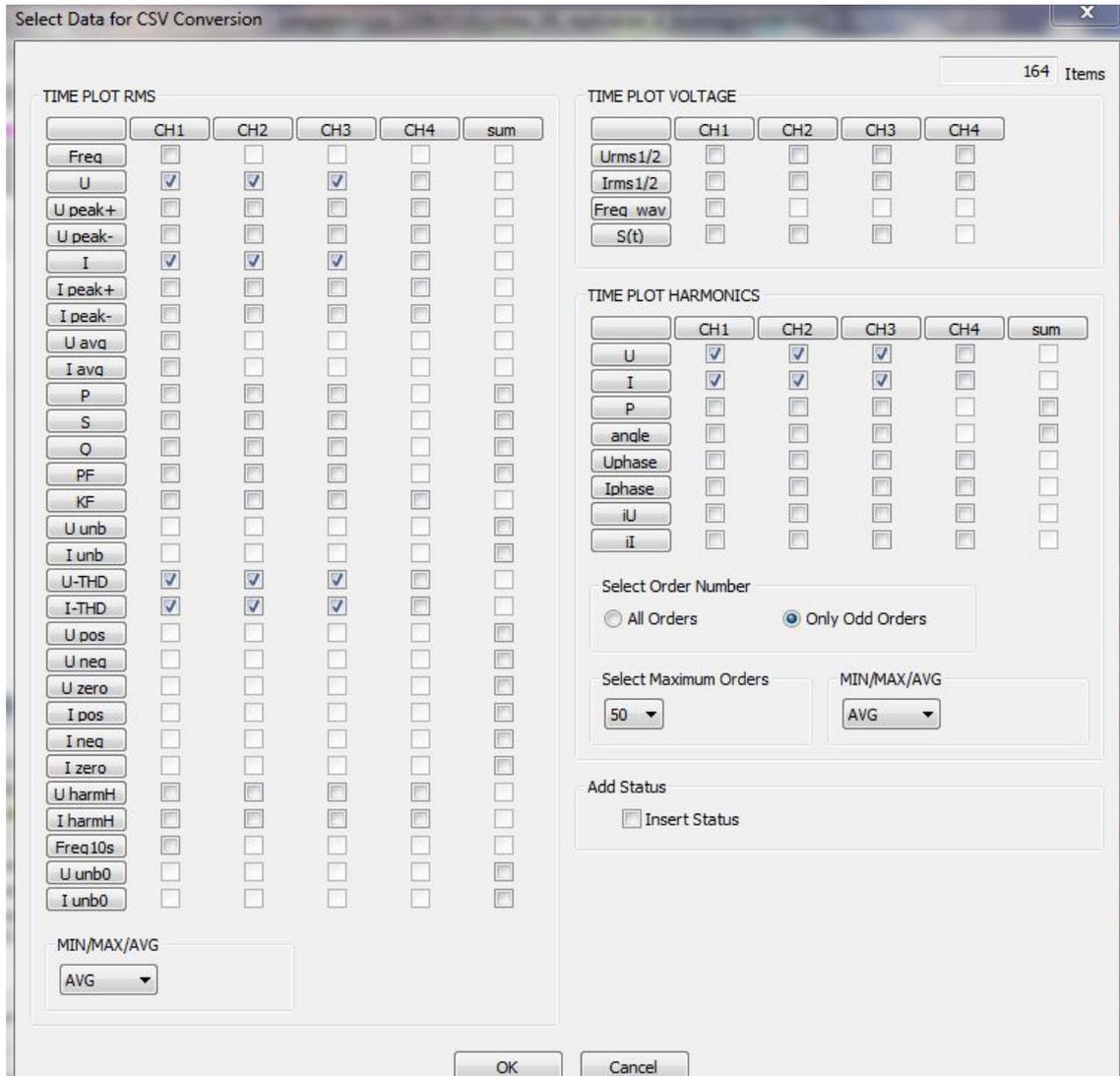
É necessário primeiro, desenvolver uma compreensão do domínio do aplicativo e dos antecedentes relevantes.

O período de coleta foi de uma semana de duração, as coletas dos dados escolhidos foram nos seguintes pontos: um na cabine de medição geral, e um em cada um dos alimentadores AL-01, AL-02, AL-03 e AL-04, os quais estão distribuídos entre os Campi Básico 1 e 2, Profissional e Saúde. Nestes pontos, foram instalados 5 medidores de qualidade de energia HIOKI PW3198.

6.3.2 Etapa de Seleção de Dados

Nesta etapa os dados foram exportados dos analisadores de QEE, com ajuda de um software de exportação, disponibilizado pelo fabricante do analisador conforme figura 13.

Figura 13 - Seleção dos Dados do Analisador para Exportação.



Fonte: Software Hioki, (2019).

A figura 13 apresenta os itens que foram selecionado/marcados para exportação: THD de tensão, corrente e harmônicos de ordem 1 a 49, as quais foram analisados nas fases seguintes.

O formato escolhido para exportação foi o *CSV* (**Comma-separated values**), ou seja, os dados foram disponibilizados no formato lido por planilhas de cálculos. Das medições realizadas no Campus Universitário, foram exportados 5 arquivos no formato *CSV*, um referente a cada uma das áreas analisadas. O analisador coleta uma grande variedade de dados,

neste momento foi necessário que o especialista do domínio analisasse os dados para definir quais seriam utilizados nas avaliações do conteúdo harmônico existente em cada um dos pontos analisados, foram definidos 4 tipos de dados (datas, horas, números reais e caracteres) os quais possuíam mais relevância para as análises.

Tabela 13 - Especificação técnica dos dados dos arquivos CSV.

Nome de Arquivos	Total de colunas	Total de Linhas	Tipos de dados encontrados arquivos			
Básico 1	164	11.841	Números reais	Caracteres	Datas	Horas
Básico 2	164	11.844				
Entrada Geral UFPA	164	11.818				
Profissional	164	11.841				
Saúde	164	11.844				

Fonte: Autor, (2019).

Conforme definido pelo especialista do domínio, a partir dos dados selecionados foram criadas as estruturas de cada uma das tabelas no banco de dados que receberam os dados oriundos dos arquivos CSV, isso se faz necessário em virtude de existirem uma variedade muito grande de colunas e linhas, então, optou-se por criar as tabelas com os seguintes dados: id, data, hora, THDu de cada fase, TDHi de cada fase, harmônicos de tensão de ordem 1° até 49°, harmônicos de corrente de ordem 1° até 49° dos respectivos campi. A tabela 13, apresenta os tipos de dados que foram identificados nos arquivos CSV.

Optou-se por usar um SGBD por este apresentar as seguintes características:

- Controlar redundância;
- Compartilhamento de dados;
- Independência de dados;
- Segurança;
- Backup e recuperação a falhas;
- Forçar restrições de integridade;
- Aumentar a produtividade e disponibilidade;
- Flexibilidade, padronização.

A tabela 14 apresenta os campos que foram identificados nos arquivos CSV.

Tabela 14 - Estrutura dos campos das tabelas no SGDB.

Campos	Descrição	Tipo de dados
Id	Ordem do registro	Numérico
Data	Data da coleta	Data
Hora	Hora da coleta	Hora
Avgurms1, 2 e 3	Tensão das fases A, B e C	Numérico científico
Avgirms1, 2 e 3	Corrente das fases A, B e C	
Avguthd1, 2 e 3	THD de tensão das fases A, B e C	
Avgithd1, 2 e 3	THD de correntes das fases A, B e C	
Avgu1_1, Avgu1_3, Avgu1_5, Avgu1_7, Avgu1_9, Avgu1_11, Avgu1_13, Avgu1_15 até Avgu1_49	Harmônicas de tensão de ordem 1° até 49° da fase A	
Avgu2_1, Avgu2_3, Avgu2_5, Avgu2_7, Avgu2_9, Avgu2_11, Avgu2_13, Avgu2_15 até Avgu2_49	Harmônicas de tensão de ordem 1° até 49° da fase B	
Avgu3_1, Avgu3_3, Avgu3_5, Avgu3_7, Avgu3_9, Avgu3_11, Avgu3_13, Avgu3_15 até Avgu3_49	Harmônicas de tensão de ordem 1° até 49° da fase C	
Avgi1_1, Avgi1_3, Avgi1_5, Avgi1_7, Avgi1_9, Avgi1_11, Avgi1_13, Avgi1_15 até Avgi1_49	Harmônicas de corrente de ordem 1° até 49° da fase A	
Avgi2_1, Avgi2_3, Avgi2_5, Avgi2_7, Avgi2_9, Avgi2_11, Avgi2_13, Avgi2_15 até Avgi2_49	Harmônicas de corrente de ordem 1° até 49° da fase B	
Avgi3_1, Avgi3_3, Avgi3_5, Avgi3_7, Avgi3_9, Avgi3_11, Avgi3_13, Avgi3_15 até Avgi3_49	Harmônicas de corrente de ordem 1° até 49° da fase C	

Fonte: Autor, (2019).

Posteriormente a definição de nomes dos campos e seus tipos, as tabelas foram criadas no SGDB. O passo seguinte foi a importação dos arquivos csv para as tabelas do banco de dados, facilitando assim, o processo de análise, visualização, apresentação e disponibilidade dos dados para as próximas fases do KDD.

6.3.3 Etapa de Pré-Processamento

Nesta fase, o objetivo é selecionar os dados mais relevantes em relação à tarefa de aprendizagem, a confiabilidade dos dados foi aprimorada. Inclui-se remoção de dados, como o manuseio de valores em falta e a remoção de ruído ou *outliers*, para que os mesmos possam ser representados e utilizados na construção de modelos confiáveis, tornando-se assim, mais fáceis e rápidos os possíveis ajustes que poderão ocorrer durante o processo.

6.3.4 A Fase de Transformação

Aplicou-se a formatação de dados, isto se faz necessário, devido aos dados encontrarem-se em formatos e unidades diferentes, impossibilitando a tarefa de reconhecimento dos dados por parte do algoritmo no processo de KDD. Então foi necessário

a execução dos processos de formatações, conforme pode ser observado na tabela 15 os campos, os formatos originais, bem como os novos formatos e a redução de quantidade de casas decimais.

Tabela 15 - Transformação dos dados.

Campos identificados	Formatos Originais	Formatação Aplicada	Observações
Avgurms1, 2 e 3	0,12373E+0	0.124	Troca de padrão de notação científica para o padrão decimal e redução para três casas decimais.
Avgirms1, 2 e 3			
Avguthd1, 2 e 3			
Avgithd1, 2 e 3			
Avgu1_1, Avgu1_3, Avgu1_5, Avgu1_7, Avgu1_9, Avgu1_11 até Avgu1_49			
Avgu2_1, Avgu2_3, Avgu2_5, Avgu2_7, Avgu2_9, Avgu2_11 até Avgu2_49			
Avgu3_1, Avgu3_3, Avgu3_5, Avgu3_7, Avgu3_9, Avgu3_11 até Avgu3_49			
Avgi1_1, Avgi1_3, Avgi1_5, Avgi1_7, Avgi1_9, Avgi1_11 até Avgi1_49			
Avgi2_1, Avgi2_3, Avgi2_5, Avgi2_7, Avgi2_9, Avgi2_11 até Avgi2_49			
Avgi3_1, Avgi3_3, Avgi3_5, Avgi3_7, Avgi3_9, Avgi3_11 até Avgi3_49			

Fonte: Autor, (2019).

Utilizando-se o conhecimento de um especialista no domínio, foi necessário realizar as transformações e alteração de nomes das colunas “*tuplas ou campos*”, presentes em cada uma das tabelas. Estas mudanças justificam-se, devido as tabelas originais de cada processo monitorados possuírem 164 colunas/tuplas e até 11.844 linhas, com os dados de harmônicas de 1º até 49º ordem, datas, horários, entre outros atributos.

Aplicou-se a redução da dimensionalidade, isto se faz necessário, devido a nem todos os dados apresentarem informações relevantes tal fato impossibilitaria a tarefa de reconhecimento de padrões por parte do algoritmo no processo de KDD.

Com a orientação do especialista decidiu-se criar três novas tabelas com os campos descritos nas tabelas 16, 17 e 18 referente as fases A, B e C da rede elétrica e considerar as harmônicas de ordem 3º, 5º, 7º, 9º e 11º. De acordo com as análises do especialista no domínio as demais colunas, possuíamos uma grande quantidade de dados com níveis harmônicos muito baixos e de pouca relevância para as análises. Após a execução destes processos os dados serão dimensionados de forma a reduzir as análises executadas pelo algoritmo, contribuindo assim, para melhor o desempenho da execução das fases futuras.

Tabela 16 - Nomenclaturas dos atributos utilizados nas reduções da fase A.

itens	Nome	Descrição dos dados da fase A	Tipos de dados	Decimais
1	Id	Índice de organização entrada de dados na tabela	Numero	Não possui
2	thd_fase_a_ent	Distorção harmônica total	Numero	3 casas
3	h3_i_basico1	Distorção harmônica de corrente 3° ord. basico1		
4	h5_i_basico1	Distorção harmônica de corrente 5° ord. basico1		
5	h7_i_basico1	Distorção harmônica de corrente 7° ord. basico1		
6	h9_i_basico1	Distorção harmônica de corrente 9° ord. basico1		
7	h11_i_basico1	Distorção harmônica de corrente 11° ord. basico1		
8	h3_i_basico2	Distorção harmônica de corrente 3° ord. basico2		
9	h5_i_basico2	Distorção harmônica de corrente 5° ord. basico2		
10	h7_i_basico2	Distorção harmônica de corrente 7° ord. basico2		
11	h9_i_basico2	Distorção harmônica de corrente 9° ord. basico2		
12	h11_i_basico2	Distorção harmônica de corrente 11° ord. basico2		
13	h3_i_profissional	Distorção harmônica de corrente 3° ord. profissional		
14	h5_i_profissional	Distorção harmônica de corrente 5° ord. profissional		
15	h7_i_profissional	Distorção harmônica de corrente 7° ord. profissional		
16	h9_i_profissional	Distorção harmônica de corrente 9° ord. profissional		
17	h11_i_profissional	Distorção harmônica de corrente 9° ord. profissional		
18	h3_i_saude	Distorção harmônica de corrente 3° ord. saúde		
19	h5_i_saude	Distorção harmônica de corrente 5° ord. saúde		
20	h7_i_saude	Distorção harmônica de corrente 7° ord. saúde		
21	h9_i_saude	Distorção harmônica de corrente 9° ord. saúde		
22	h11_i_saude	Distorção harmônica de corrente 11° ord. saúde		

Fonte: Autor, (2019).

Os atributos “campos” descritos nas tabelas 16, 17 e 18 são utilizados para armazenar os dados oriundos dos alimentadores analisados nas fases A, B e C do sistema elétrico que alimenta a Universidade Federal do Pará. Os nomes de cada dado representam a origem do alimentador, os tipos de dados e suas quantidades de casas decimais utilizados nas análises.

Tabela 17 - Nomenclaturas dos atributos utilizados nas reduções da fase B.

itens	Nome	Descrição dos dados da fase B	Tipos de dados	Decimais
1	Id	Índice de organização entrada de dados na tabela	Numero	Não possui
2	thd_fase_b_ent	Distorção harmônica total	Numero	3 casas
3	h3_i_basico1	Distorção harmônica de corrente 3º ord. basico1		
4	h5_i_basico1	Distorção harmônica de corrente 5º ord. basico1		
5	h7_i_basico1	Distorção harmônica de corrente 7º ord. basico1		
6	h9_i_basico1	Distorção harmônica de corrente 9º ord. basico1		
7	h11_i_basico1	Distorção harmônica de corrente 11º ord. basico1		
8	h3_i_basico2	Distorção harmônica de corrente 3º ord. basico2		
9	h5_i_basico2	Distorção harmônica de corrente 5º ord. basico2		
10	h7_i_basico2	Distorção harmônica de corrente 7º ord. basico2		
11	h9_i_basico2	Distorção harmônica de corrente 9º ord. basico2		
12	h11_i_basico2	Distorção harmônica de corrente 11º ord. basico2		
13	h3_i_profissional	Distorção harmônica de corrente 3º ord. profissional		
14	h5_i_profissional	Distorção harmônica de corrente 5º ord. profissional		
15	h7_i_profissional	Distorção harmônica de corrente 7º ord. profissional		
16	h9_i_profissional	Distorção harmônica de corrente 9º ord. profissional		
17	h11_i_profissional	Distorção harmônica de corrente 9º ord. profissional		
18	h3_i_saude	Distorção harmônica de corrente 3º ord. saúde		
19	h5_i_saude	Distorção harmônica de corrente 5º ord. saúde		
20	h7_i_saude	Distorção harmônica de corrente 7º ord. saúde		
21	h9_i_saude	Distorção harmônica de corrente 9º ord. saúde		
22	h11_i_saude	Distorção harmônica de corrente 11º ord. saúde		

Fonte: Autor, (2019).

Tabela 18 - Nomenclaturas dos atributos utilizados nas reduções da fase C.

itens	Nome	Descrição dos dados da fase C	Tipos de dados	Decimais
1	Id	Índice de organização entrada de dados na tabela	Numero	Não possui
2	thd_fase_c_ent	Distorção harmônica total	Numero	3 casas
3	h3_i_basico1	Distorção harmônica de corrente 3° ord. basico1		
4	h5_i_basico1	Distorção harmônica de corrente 5° ord. basico1		
5	h7_i_basico1	Distorção harmônica de corrente 7° ord. basico1		
6	h9_i_basico1	Distorção harmônica de corrente 9° ord. basico1		
7	h11_i_basico1	Distorção harmônica de corrente 11° ord. basico1		
8	h3_i_basico2	Distorção harmônica de corrente 3° ord. basico2		
9	h5_i_basico2	Distorção harmônica de corrente 5° ord. basico2		
10	h7_i_basico2	Distorção harmônica de corrente 7° ord. basico2		
11	h9_i_basico2	Distorção harmônica de corrente 9° ord. basico2		
12	h11_i_basico2	Distorção harmônica de corrente 11° ord. basico2		
13	h3_i_profissional	Distorção harmônica de corrente 3° ord. profissional		
14	h5_i_profissional	Distorção harmônica de corrente 5° ord. profissional		
15	h7_i_profissional	Distorção harmônica de corrente 7° ord. profissional		
16	h9_i_profissional	Distorção harmônica de corrente 9° ord. profissional		
17	h11_i_profissional	Distorção harmônica de corrente 9° ord. profissional		
18	h3_i_saude	Distorção harmônica de corrente 3° ord. saúde		
19	h5_i_saude	Distorção harmônica de corrente 5° ord. saúde		
20	h7_i_saude	Distorção harmônica de corrente 7° ord. saúde		
21	h9_i_saude	Distorção harmônica de corrente 9° ord. saúde		
22	h11_i_saude	Distorção harmônica de corrente 11° ord. saúde		

Fonte: Autor, (2019).

Uma vez que as definições de redução dos dados foram executadas, iniciou-se o processo de alimentar as três novas tabelas, executando-se o sincronismo de dados, a partir das 5 tabelas origens (entrada_tensão, basico1, basico2, profissional, saúde), para tanto executou-se comandos SQL “*Structured Query Language*”. Utilizou-se discretização não-supervizada, com o objetivo de converter um atributo numérico (contínuo) em discreto (categorizado). Verificou-se que a criação de faixas de igual frequência pode obter excelentes resultados, pelo menos em conjunto com o esquema de aprendizagem Naive Bayes, quando o número de compartimentos é escolhido de uma forma dependente dos dados, fixando-o à raiz quadrada do número dos casos (WITTEN, FRANK et al., 2016).

A figura 14 apresenta os cinco intervalos discretizados por frequência da 7^o harmônica do básico 1.

Figura 14 - Dados discretizados harmônica 7 básico 1.

Selected attribute			
Name: h7_i_basico1		Type: Nominal	
Missing: 0 (0%)		Distinct: 5	
		Unique: 0 (0%)	
No.	Label	Count	Weight
1	'(-inf-0.585]'	1976	1976.0
2	'(0.585-0.885]'	1971	1971.0
3	'(0.885-0.995]'	1961	1961.0
4	'(0.995-1.185]'	2064	2064.0
5	'(1.185-inf)'	2028	2028.0

Fonte: O Autor, (2019).

6.3.5 Etapa de Objetivos do Processo KDD

A determinação dos objetivos do processo KDD, é o primeiro passo para um método particular de mineração de dados. A classificação é uma forma de análise de dados que extrai modelos que descrevem classes importantes de dados. Tais modelos, chamados classificadores, predizem rótulos de classe categóricos (discretos, não ordenados), por tais motivos este foi o método escolhido.

6.3.6 Etapa de análise exploratória e seleção de modelo e hipótese

Nesta etapa escolheu-se o algoritmo de mineração de dados e os métodos de seleção a serem usados para pesquisar padrões de dados. Utilizou-se a técnica de classificação Naive Bayes, um classificador de Naive Bayes é um simples classificador probabilístico baseado na aplicação do teorema de Bayes, onde cada característica é assumida como independente da classe. O classificador Naive Bayes assume que o efeito de um valor de uma variável em uma determinada classe é independente dos valores de outras variáveis. Esse pressuposto é chamado de independência de classe. É feito para simplificar a computação e, nesse sentido, é considerado ingênuo.

6.3.7 Etapa de Mineração de dados

Finalmente, a aplicação do algoritmo de mineração de dados é alcançada. Nesta etapa, talvez seja necessário empregar o algoritmo várias vezes até obter um resultado satisfatório. Considerada como a fase ou etapa de maior importância, nesta fase será utilizada a técnica de classificação de Naive Bayes conforme definido na etapa anterior, validação cruzada de dados, após a execução do algoritmo obteve-se os gráficos de desempenho para cada uma das áreas (básico 1, básico 2, profissional e saúde) e fases A, B e C da entrada.

Com as análises das três fases foram geradas seis tabelas, das quais três com as matrizes de confusão e três com as análises de acurácia do algoritmo detalhadas, além dos gráficos de Pareto para cada uma das áreas e fases. Nas tabelas 19, 20 e 21 estão todos os dados referentes as matrizes de confusão, já nas tabelas 22, 23 e 24 estão descritos os dados sobre a precisão ou acurácia obtidas pelo algoritmo, os dados descritos nas tabelas são utilizados nas análises de desempenho do algoritmo referentes a cada uma das três fases.

A matriz de confusão disponibiliza uma avaliação de medida de desempenho do modelo de classificação, ao demonstrar a quantidade de classificações corretas versus as classificações preditas para cada classe, sobre um conjunto de exemplos.

- A **matriz de confusão** de uma hipótese h oferece uma medida efetiva do modelo de classificação, ao mostrar o número de classificações corretas *versus* as classificações preditas para cada classe, sobre um conjunto de exemplos T
- O número de acertos, para cada classe, se localiza na diagonal principal $M(C_i, C_i)$ da matriz
- Os demais elementos $M(C_i, C_j)$, para $i \neq j$, representam erros na classificação
- A matriz de confusão de um classificador ideal possui todos esses elementos iguais a zero uma vez que ele não comete erros

Tabela 19 - Matriz de Confusão para duas Classes.

CLASSE	Predita C ₊	Predita C ₋	Taxa de Erro da Classe	Taxa de Erro Total
Verdadeira C ₊	TP	FP	FN/(TP+FN)	FP+FN/n
Verdadeira C ₋	FN	TN	FP/(FP+FN)	

Fonte o Autor (2019)

Tabela 20 - Termos utilizados no desempenho do classificador.

Termos	Descrição
TP = True Positive (Verdadeiros positivos)	Total de casos verdadeiramente classificados pelo algoritmo como positivos.
FP = False Positive (Falsos Positivo)	Total de casos classificados pelo algoritmo como positivos que são falsos.
FN = Falsos Negativos (False Negative)	Total de casos classificados pelo algoritmo como negativos que não são negativos.
TN = True Negative (Verdadeiros Negativos)	Total de casos classificados pelo algoritmo como verdadeiros negativos.
$n = (TP + FP + FN + TN)$	Somatório
TPR = TP Rate (Taxa de verdadeiros)	Porcentagem de amostras corretamente classificadas como positivas dentre todas as positivas reais
FPR = FP Rate (Taxa de Falsos positivos.)	Porcentagem de amostras erroneamente classificadas como positivas dentre todas as negativas reais
Precision = Precisão	Porcentagem de amostras positivas classificadas corretamente sobre o total de amostras classificadas como positivas
Curva ROC	É um gráfico da verdadeira taxa positiva versus taxa de falso positivo

Fonte o Autor (2019)

- Acurácia: porcentagem de amostras positivas e negativas classificadas corretamente sobre a soma de amostras positivas e negativas.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (13)$$

- Sensitividade (Recall): porcentagem de amostras positivas classificadas corretamente sobre o total de amostras positivas

$$\text{Sensitivity} = \frac{TP}{TP + FN} = \frac{TP}{\text{Positive}} \quad (14)$$

- Precisão: porcentagem de amostras positivas classificadas corretamente sobre o total de amostras classificadas como positivas

$$\text{Precision} = \frac{TP}{TP + FP} \quad (15)$$

- Especificidade: porcentagem de amostras negativas identificadas corretamente sobre o total de amostras negativas

$$\text{Specificity} = \frac{TN}{TN + FP} = \frac{TN}{\text{Negative}} \quad (16)$$

- Curva ROC (*Receiver Operating Characteristic*)
 - Cada ponto na curva corresponde a um dos modelos induzidos pelo classificador
 - Um bom modelo deve estar localizado próximo do ponto (0,1)
 - Modelos localizados na diagonal são modelos aleatórios $TPR = FPR$
 - Modelos localizados acima da diagonal são melhores do que modelos abaixo da diagonal.
- Análise da curva ROC
 - Ponto (0,1) é o classificador perfeito: classifica todas as amostras positivas e negativas corretamente. $FPR=0$ e $TPR=1$.
 - O ponto (0,0) representa um classificador que classifica todas as amostras como negativas, enquanto o ponto (1,1) corresponde a um classificador que classifica todas as amostras como positivas.
 - O ponto (1,0) é o classificador que classifica incorretamente todas as amostras.
 - Em muitos casos, os classificadores possuem um parâmetro que pode ser ajustado para aumentar TP aumentando também FP. Cada parâmetro fornece um par (FP, TP). Um classificador não-paramétrico é representado por um único ponto na curva ROC.

Apresentamos as análises detalhadas da acurácia do Naive Bayes, para tanto utilizou-se as curvas ROC de cada classe, atentando-se para os valores da Taxa de Verdadeiros Positivos (*TP Rate*), Taxa de Falsos Positivos (*FP Rate*), Precisão (*Precision*), Sensitividade (*Recall*) e Área da Curva ROC (*ROC Area*).

- Dados referentes Naive Bayes Fase A:

Tabela 21 - Matriz confusão da fase A.

a	b	c	d	e	← Classificado como
1895	86	11	0	0	a = '(-inf-1.165]'
393	1095	214	258	6	b = '(1.165-1.385]'
74	343	882	529	254	c = '(1.385-1.595]'
0	49	437	973	568	d = '(1.595-1.735]'
0	4	115	462	1352	e = '(1.735-inf)'

Fonte: Autor, (2019).

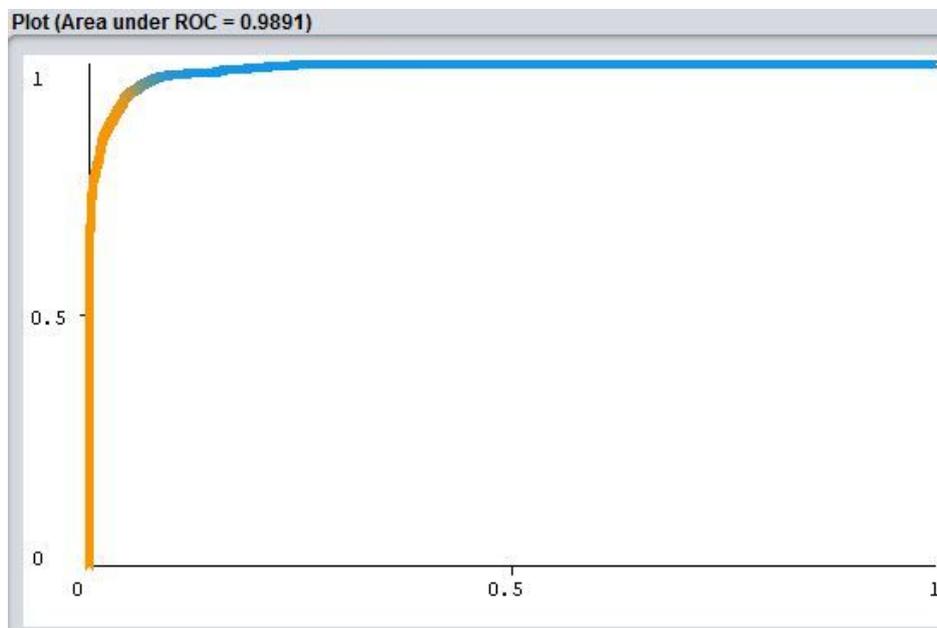
Tabela 22 - Acurácia da Fase A.

TP Taxa	FP Taxa	Precisão	Recall	ROC Área	Classes
0,951	0,058	0,802	0,951	0,989	'(-inf-1.165]'
0,557	0,060	0,694	0,557	0,905	'(1.165-1.385]'
0,424	0,098	0,532	0,424	0,825	'(1.385-1.595]'
0,480	0,157	0,438	0,480	0,808	'(1.595-1.735]'
0,699	0,103	0,620	0,699	0,917	'(1.735-inf)'
0,620	0,095	0,616	0,620	0,888	Média ponderada.

Fonte: Autor, (2019).

O gráfico da área ROC da fase A apresenta um valor de 0.998 muito próximo a 1 que é o valor ótimo para um classificador.

Figura 15 - Gráfico ROC da fase A.



Fonte: O Autor, (2019).

- Dados referente Naive Bayes Fase B:

Tabela 23 - Matriz confusão da fase B.

a	b	c	d	e	← Classificado como
---	---	---	---	---	---------------------

1768	192	6	0	0	a = '(-inf-1.135]'
465	1021	386	185	0	b = '(1.135-1.395]'
20	341	1095	378	125	c = '(2.445-2.705]'
0	92	535	741	652	d = '(1.575-1.765]'
0	0	94	267	1637	e = '(1.765-inf)'

Fonte: O Autor, (2019).

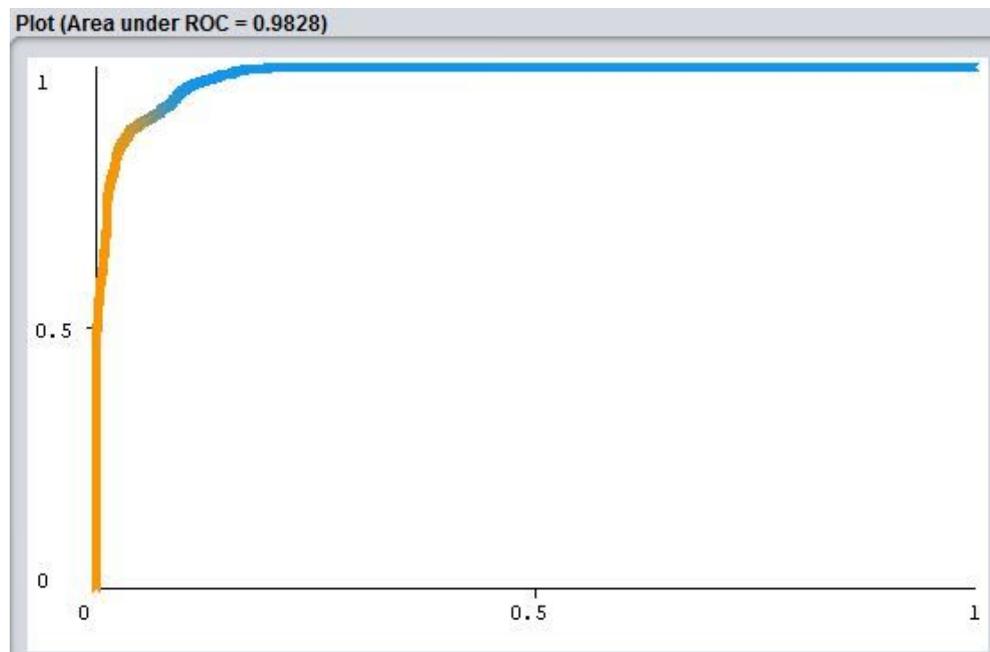
Tabela 24 - Acurácia da Fase B.

TP Taxa	FP Taxa	Precisão	Recall	ROC Área	Classes
0,899	0,060	0,785	0,899	0,983	'(-inf-1.175]'
0,496	0,079	0,620	0,496	0,876	'(1.175-1.425]'
0,559	0,127	0,517	0,559	0,845	'(1.425-1.615]'
0,367	0,104	0,472	0,367	0,821	'(1.615-1.795]'
0,819	0,097	0,678	0,819	0,949	'(1.795-inf)'
0,626	0,093	0,614	0,626	0,895	Média ponderada.

Fonte: Autor, (2019).

O gráfico da área ROC da fase B apresenta um valor de 0.982 muito próximo a 1 que é o valor ótimo para um classificador.

Figura 16 - Gráfico ROC da fase B.



Fonte: Autor, (2019).

- Dados referente Naive Bayes Fase C:

Tabela 25 - Matriz confusão da fase C.

a	b	C	d	e	← Classificado como
---	---	---	---	---	---------------------

1835	186	1	0	0	a = '(-inf-1.475]'
572	935	411	103	0	b = '(1.475-1.675]'
68	340	974	457	182	c = '(1.675-1.875]'
0	113	435	872	546	d = '(1.875-2.005]'
0	15	180	384	1391	e = '(2.005-inf)'

Fonte: Autor, (2019).

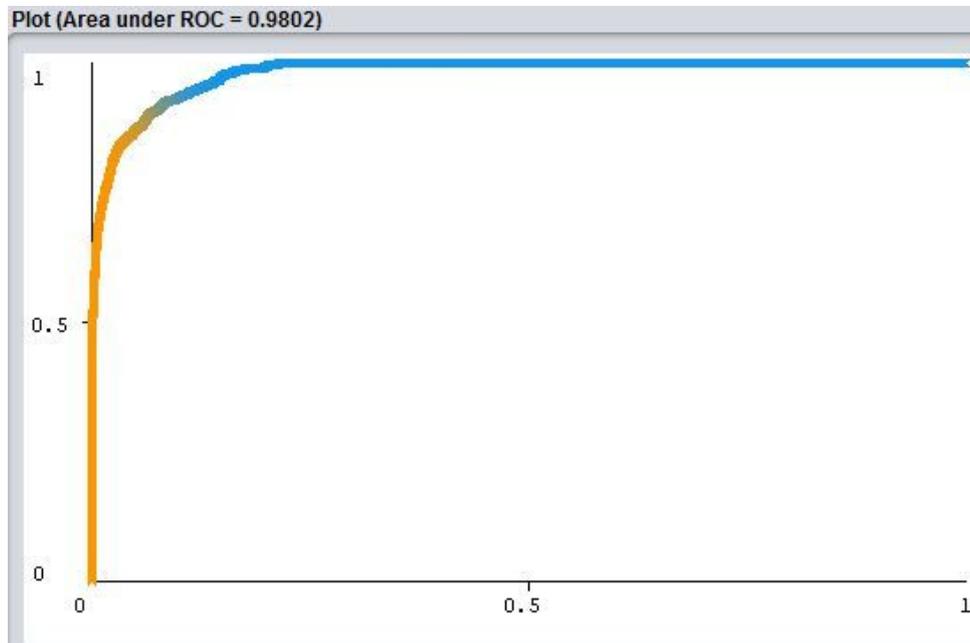
Tabela 26 - Acurácia da Fase C.

TP Taxa	FP Taxa	Precisão	Recall	ROC Área	Classes
0,908	0,075	0,753	0,908	0,980	'(-inf-1.475]'
0,479	0,109	0,534	0,479	0,847	'(1.475-1.675]'
0,443	0,106	0,506	0,443	0,815	'(1.675-1.875]'
0,450	0,125	0,479	0,450	0,828	'(1.875-2.005]'
0,723	0,087	0,665	0,723	0,935	'(2.005-inf)'
0,599	0,101	0,586	0,599	0,880	Média ponderada.

Fonte: Autor, (2019).

O gráfico da área ROC da fase B apresenta um valor de 0.980 muito próximo a 1 que é o valor ótimo para um classificador.

Figura 17 - Gráfico ROC da fase C.



Fonte: O Autor, (2019).

- Resumo dos dados analisados nas fases A, B e C:

Tabela 27 - Sumário de análises de fases - Mineração de dados.

Verificações	Fase A	Fase B	Fase C
Instâncias corretamente classificados	6197 - 61.97%	6262 - 62.62%	6007 - 60.07%
Instâncias incorretamente classificados	3803 - 38.03%	3738 - 37.38%	3993 - 39.93%
Cobertura dos casos (0,95 nível)	89.4798 %	88.386 %	90.7685 %
Rel dizer. Tamanho da região (0,95 nível)	48.3096 %	47.6551 %	51.117 %
Número total de instâncias	10000	10000	10000

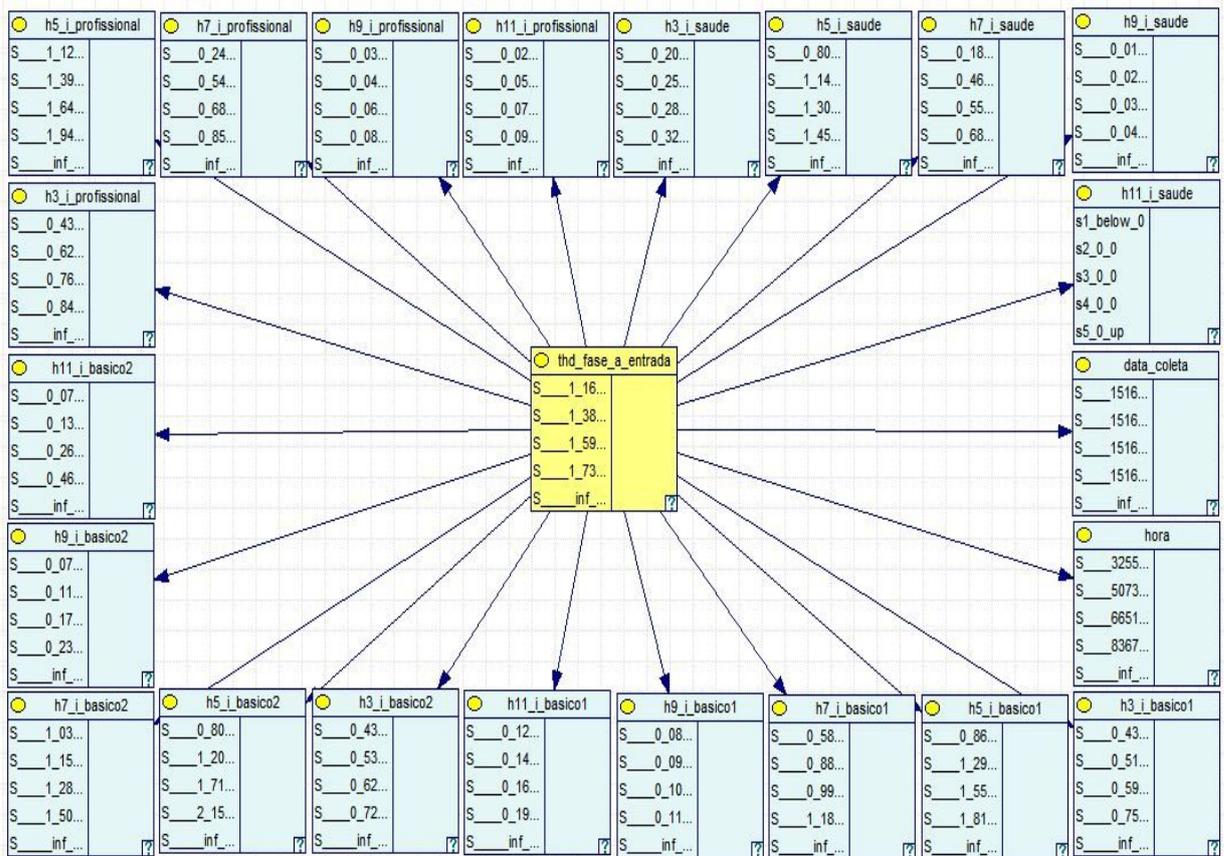
Fonte: Autor, (2019).

6.3.8 Etapa de interpretando padrões extraídos

Neste estudo, avaliou-se como as correntes harmônica de ordem 3°, 5°, 7°, 9° e 11° oriundas das áreas: Básico 1, Básico 2, Profissional e Saúde, influenciam a taxa de distorção harmônica total de tensão (THDV) da subestação de entrada. Para tal, gerou-se, por meio do classificador de Naive Bayes, três redes bayesianas uma para cada fase e os respectivos gráficos que relacionam as correntes harmônicas de ordem 3°, 5°, 7°, 9° e 11° com o THDv do secundário do transformador de entrada da subestação da UFPA campus do Guamá.

Ao se analisar as probabilidades obtidas com a criação da rede Naive Bayes para a fase A, criada com os dados provenientes da campanha de medição realizada nas instalações da UFPA, cujas grandezas registradas foram: taxa de distorção harmônica total de tensão (THDv) da cabine de medição de entrada no centro, e as correntes harmônicas de ordem 3°, 5°, 7°, 9° e 11° dos Campi: Básico 1, Básico 2, Profissional e Saúde, temos a figura 18.

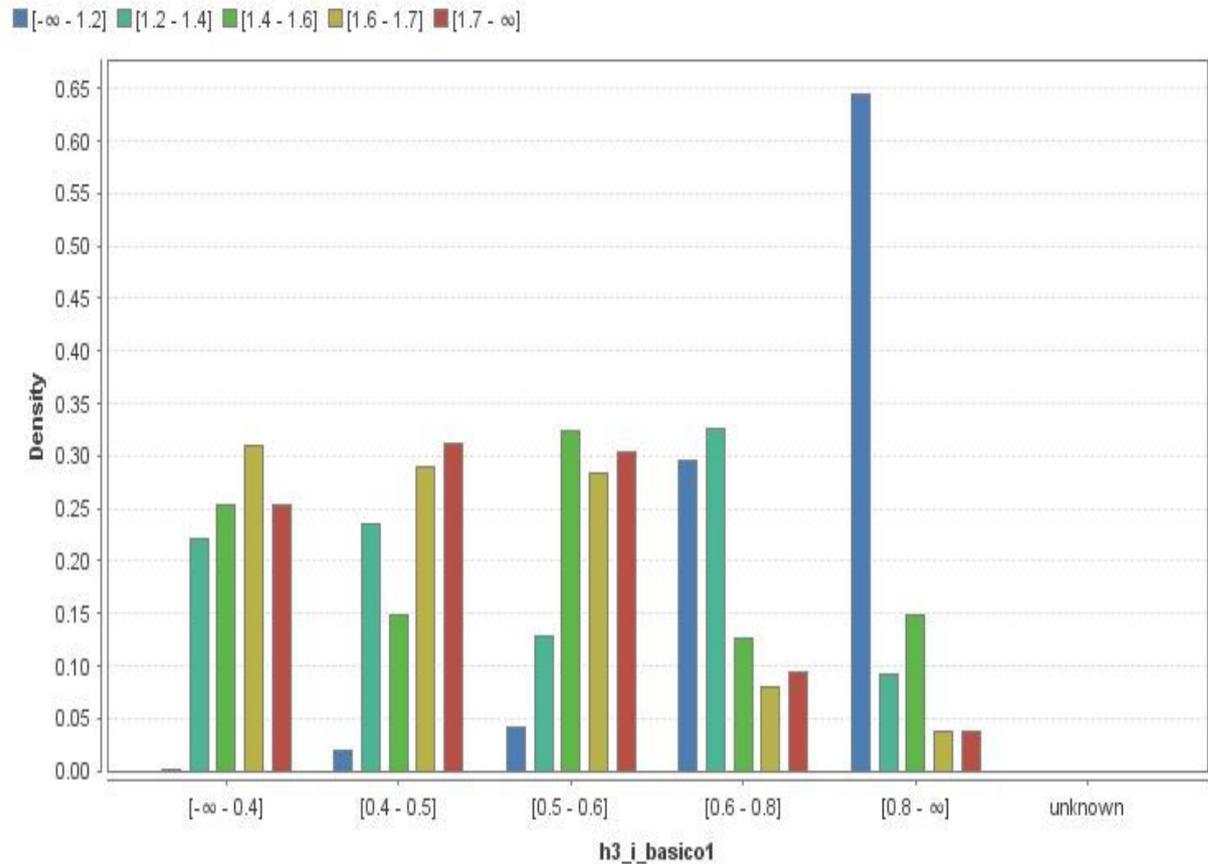
Figura 18 - Naive Bayes fase A.



Fonte: Autor, (2019).

Ao se analisar o gráfico da figura 19, que mostra a densidade de probabilidade de ocorrência dos cinco intervalos discretizados do THDv em relação a corrente do 3º harmônico do basico1, pode-se verificar que há uma probabilidade de quase 64,5% do THDv apresentar um valor igual ou maior que 1,175% quando a corrente do 3º harmônico do basico1 apresentar um valor igual ou maior que 0,8%, conforme mostra a Figura 19 a seguir.

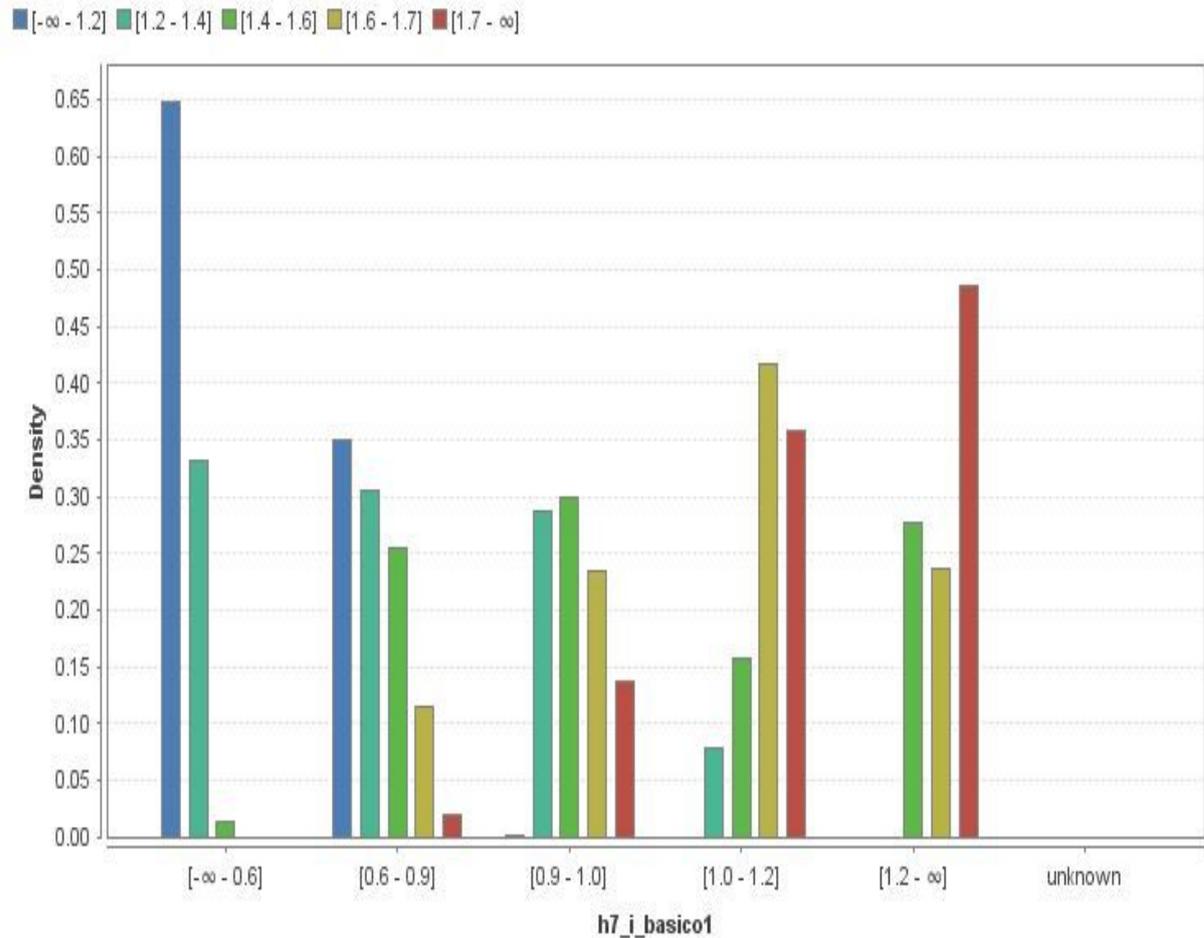
Figura 19 - Densidade de probabilidade dos intervalos discretizados do THDv em relação a corrente do 3º harmônico.



Fonte: Autor, (2019).

Ao se analisar o gráfico da figura 20, que mostra a densidade de probabilidade de ocorrência dos cinco intervalos discretizados do THDv em relação a corrente do 7º harmônico do basico1, pode-se verificar que há uma probabilidade de quase 64,8% do THDv apresentar um valor igual ou maior que 1,175% quando a corrente do 7º harmônico do basico1 apresentar um valor igual ou maior que 0,6%, conforme mostra a Figura 20 a seguir.

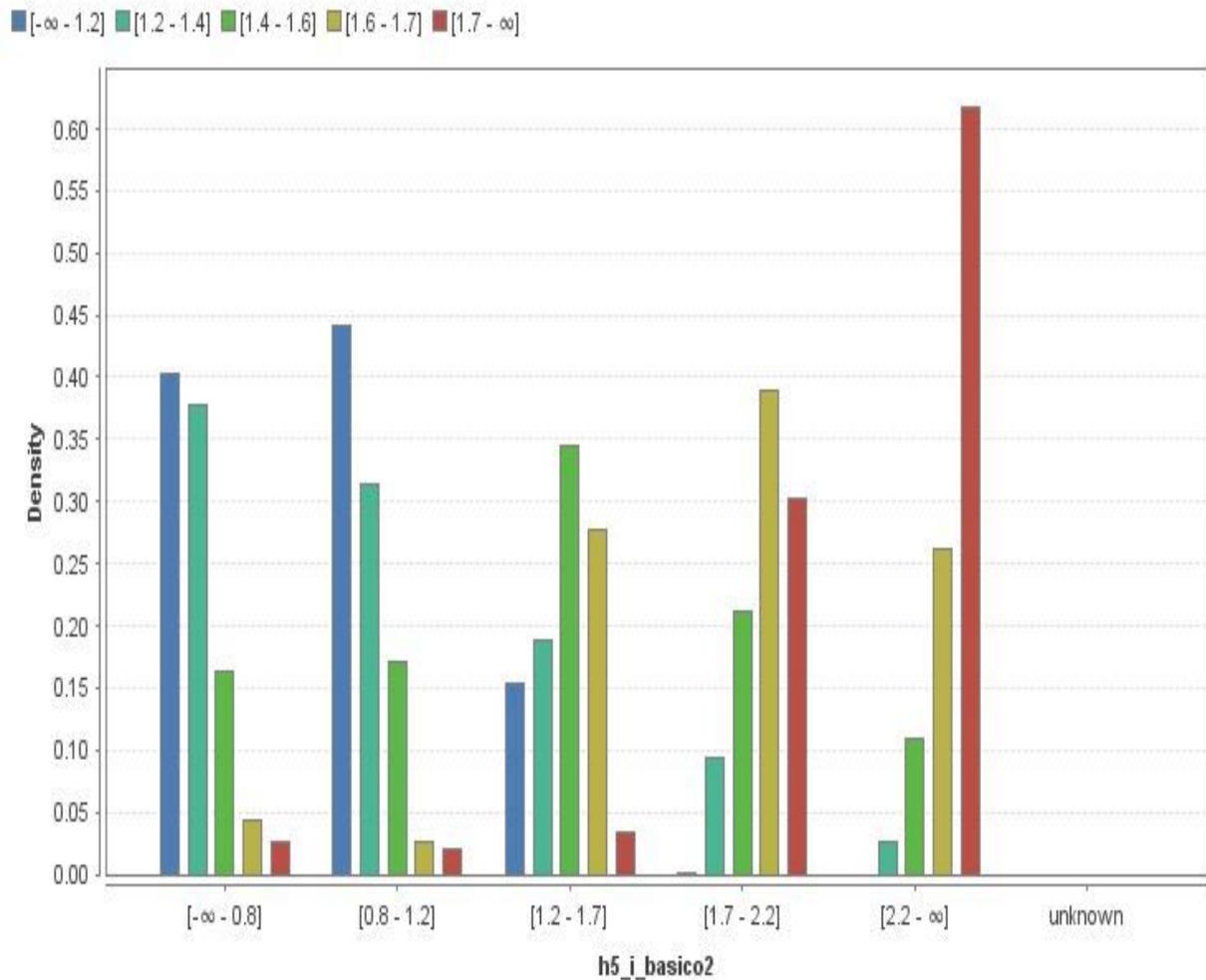
Figura 20 - Densidade de probabilidade dos intervalos discretizados do THDv em relação a corrente do 7º harmônico.



Fonte: Autor, (2019).

Ao se analisar o gráfico da figura 21, que mostra a densidade de probabilidade de ocorrência dos cinco intervalos discretizados do THDv em relação a corrente do 5º harmônico do basico1, pode-se verificar que há uma probabilidade de quase 61,8% do THDv apresentar um valor igual ou maior que 1,735% quando a corrente do 5º harmônico do basico2 apresentar um valor igual ou maior que 2,2%, conforme mostra a Figura 21 a seguir.

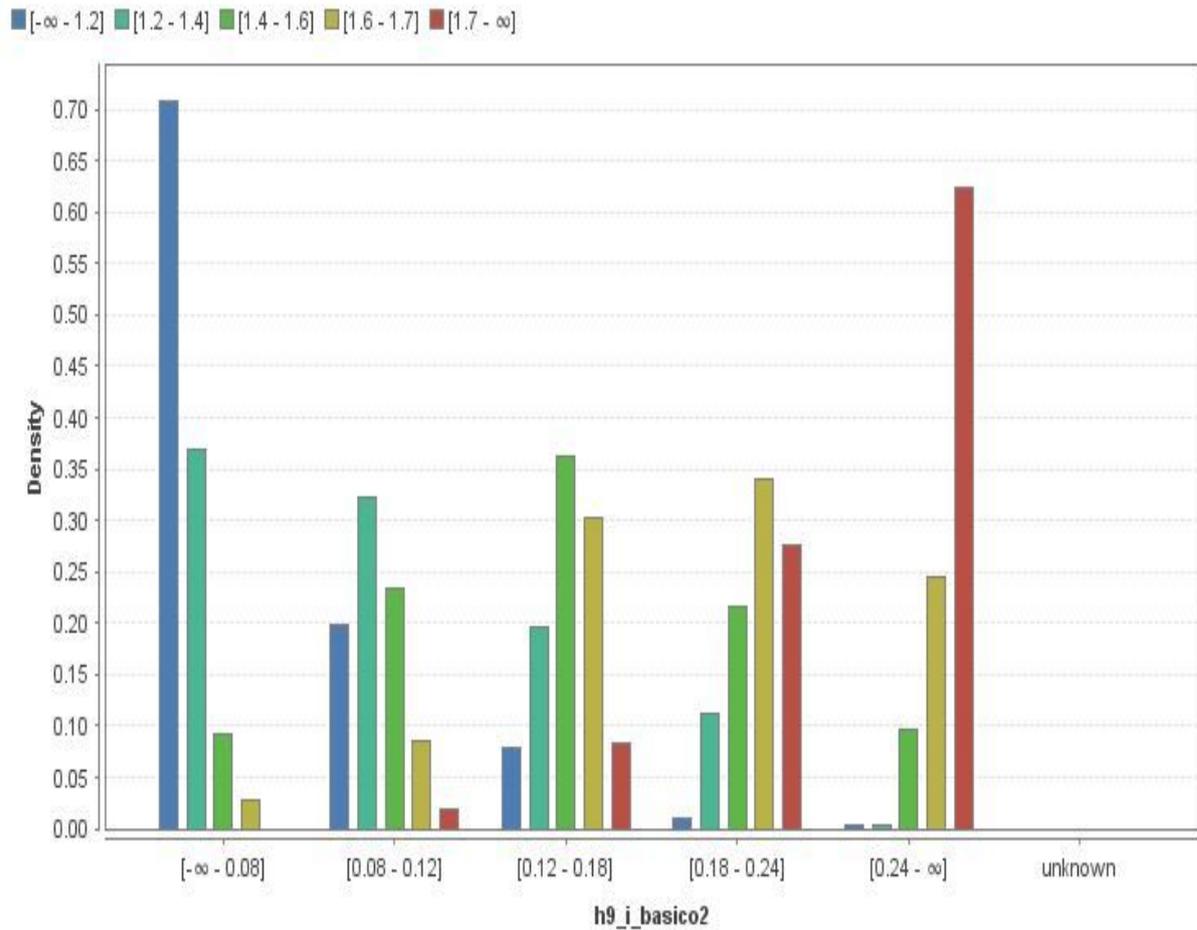
Figura 21 - Densidade de probabilidade dos intervalos discretizados do THDv em relação a corrente do 5º harmônico.



Fonte: Autor, (2019).

Ao se analisar o gráfico da figura 22, que mostra a densidade de probabilidade de ocorrência dos cinco intervalos discretizados do THDv em relação a corrente do 9º harmônico do basico2, pode-se verificar que há uma probabilidade de quase 70,9% do THDv apresentar um valor igual ou maior que 1,175% quando a corrente do 9º harmônico do basico2 apresentar um valor igual ou maior que 0,08%, conforme mostra a Figura 22 a seguir.

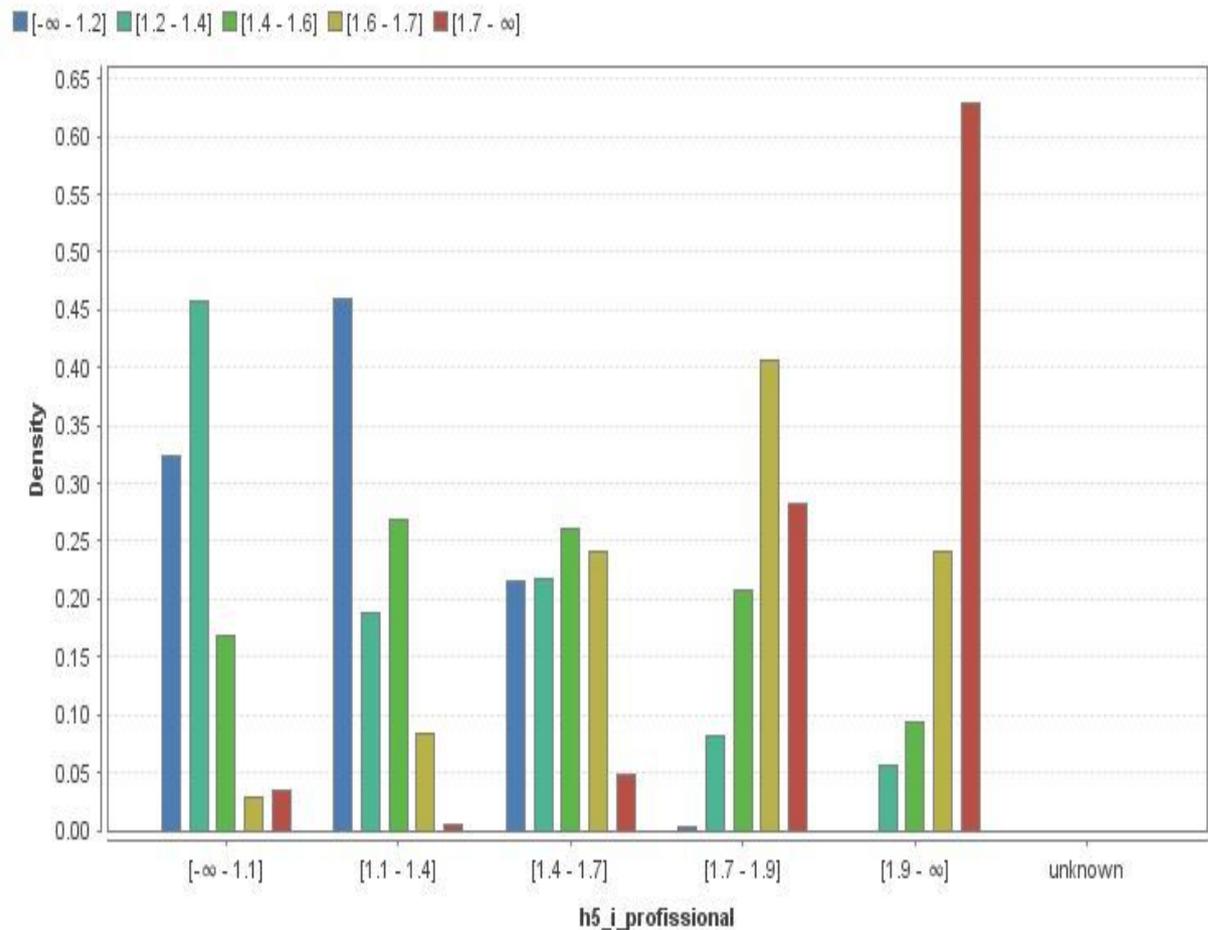
Figura 22 - Densidade de probabilidade dos intervalos discretizados do THDv em relação a corrente do 9º harmônico.



Fonte: Autor, (2019).

Ao se analisar o gráfico da figura 23, que mostra a densidade de probabilidade de ocorrência dos cinco intervalos discretizados do THDv em relação a corrente do 5º harmônico do profissional, pode-se verificar que há uma probabilidade de quase 62,9% do THDv apresentar um valor igual ou maior que 1,735% quando a corrente do 5º harmônico do profissional apresentar um valor igual ou maior que 1,9%, conforme mostra a Figura 23 a seguir.

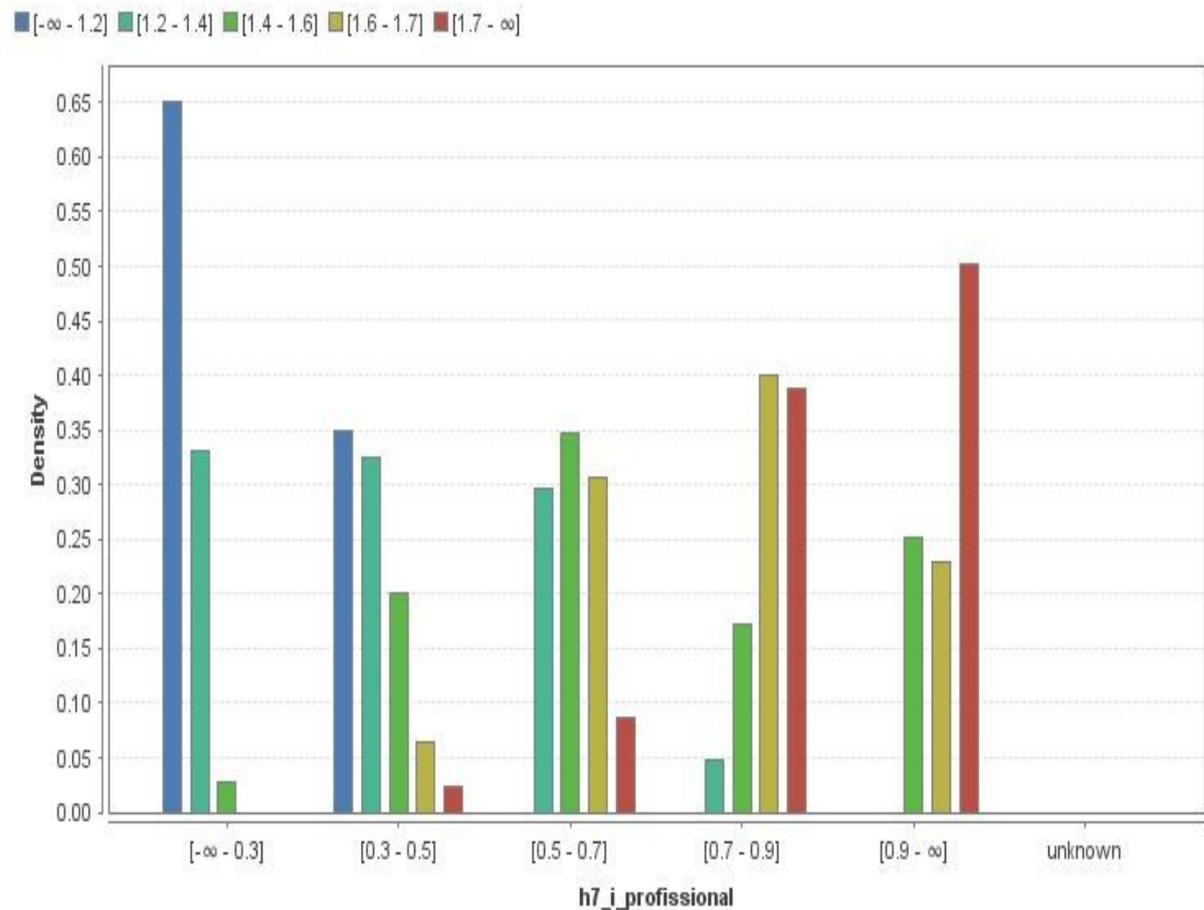
Figura 23 - Densidade de probabilidade dos intervalos discretizados do THDv em relação a corrente do 5º harmônico.



Fonte: Autor, (2019).

Ao se analisar o gráfico da figura 24, que mostra a densidade de probabilidade de ocorrência dos cinco intervalos discretizados do THDv em relação a corrente do 7º harmônico do profissional, pode-se verificar que há uma probabilidade de quase 65,1% do THDv apresentar um valor igual ou maior que 1,175% quando a corrente do 7º harmônico do profissional apresentar um valor igual ou maior que 0,3%, conforme mostra a Figura 24 a seguir.

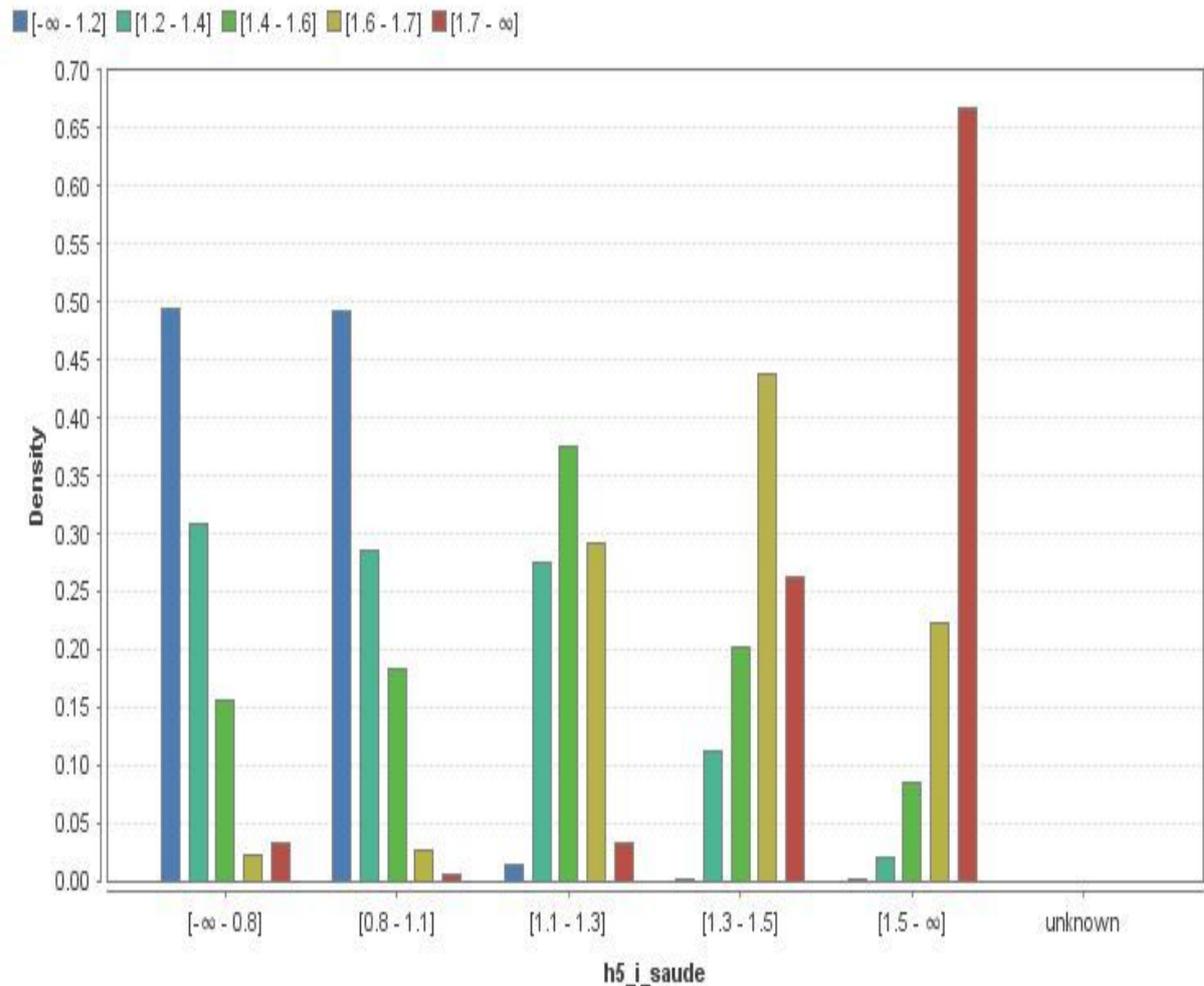
Figura 24 - Densidade de probabilidade dos intervalos discretizados do THDv em relação a corrente do 7º harmônico.



Fonte: Autor, (2019).

Ao se analisar o gráfico da figura 25, que mostra a densidade de probabilidade de ocorrência dos cinco intervalos discretizados do THDv em relação a corrente do 5º harmônico da saúde, pode-se verificar que há uma probabilidade de quase 66,8% do THDv apresentar um valor igual ou maior que 1,735% quando a corrente do 5º harmônico da saúde apresentar um valor igual ou maior que 1,5%, conforme mostra a Figura 25 a seguir.

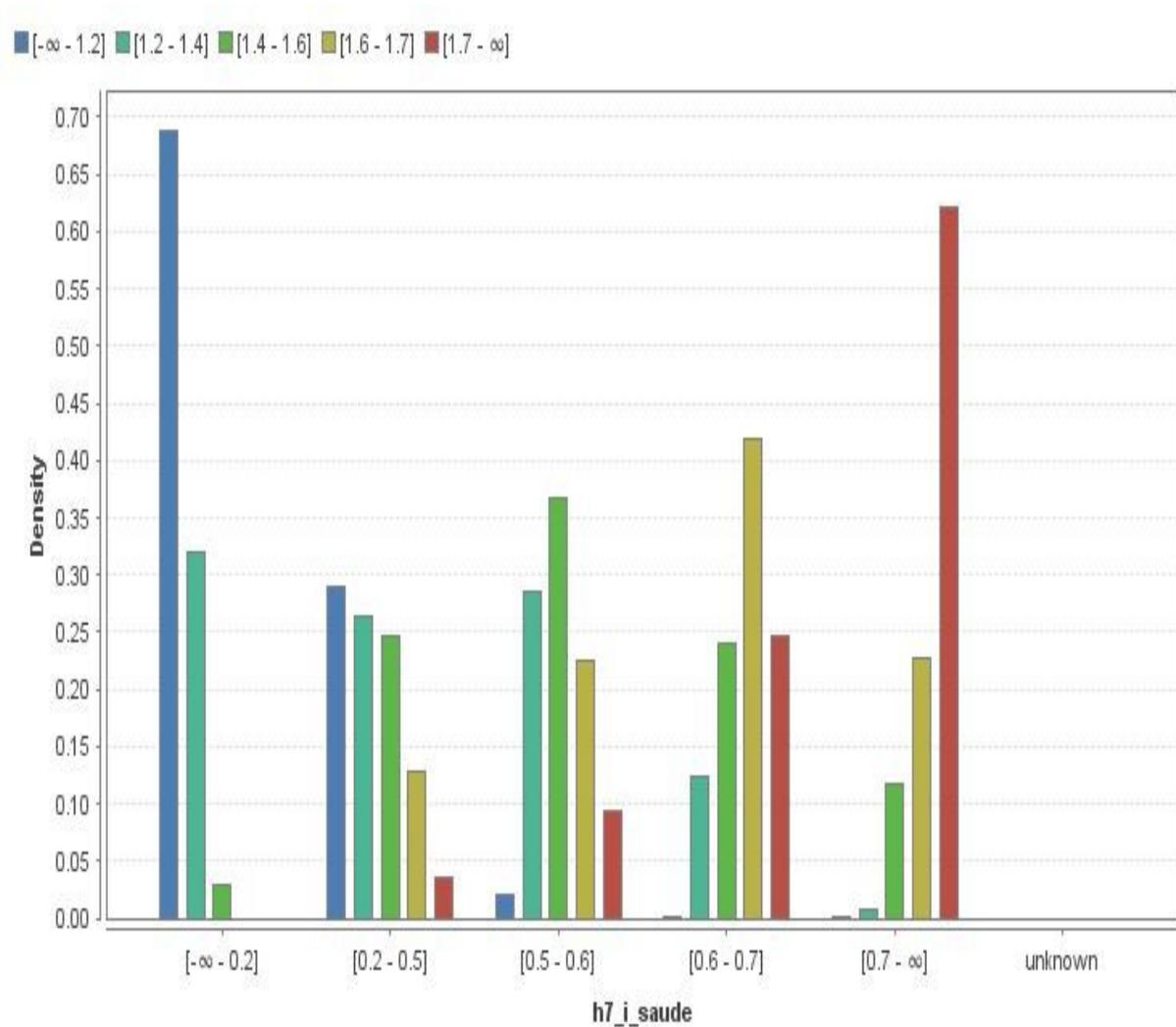
Figura 25 - Densidade de probabilidade dos intervalos discretizados do THDv em relação a corrente do 5º harmônico.



Fonte: Autor, (2019).

Ao se analisar o gráfico da figura 26, que mostra a densidade de probabilidade de ocorrência dos cinco intervalos discretizados do THDv em relação a corrente do 7º harmônico da saúde, pode-se verificar que há uma probabilidade de quase 66,9% do THDv apresentar um valor igual ou maior que 1,175% quando a corrente do 7º harmônico da saúde apresentar um valor igual ou maior que 0,2%, conforme mostra a Figura 26 a seguir.

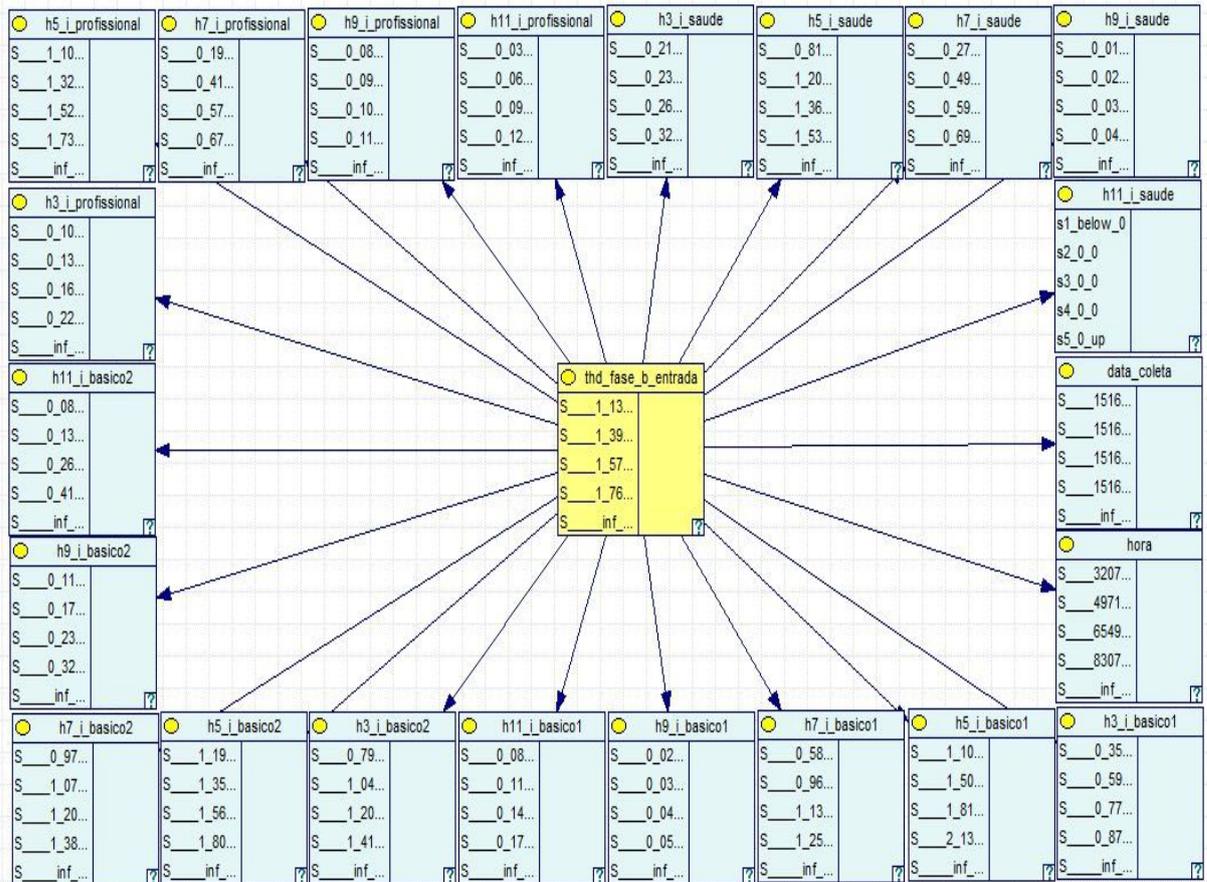
Figura 26 - Densidade de probabilidade dos intervalos discretizados do THDv em relação a corrente do 7º harmônico.



Fonte: Autor, (2019).

Ao se analisar as probabilidades obtidas com a criação da rede Naive Bayes para a fase B, criada com os dados provenientes da campanha de medição realizada nas instalações da UFPA, cujas grandezas registradas foram: taxa de distorção harmônica total de tensão (THDv) da Cabine de medição de entrada no centro, e as correntes harmônicas de ordem 3º, 5º, 7º, 9º e 11º dos Campi: Básico 1, Básico 2, Profissional e Saúde, temos a figura 27 a seguir.

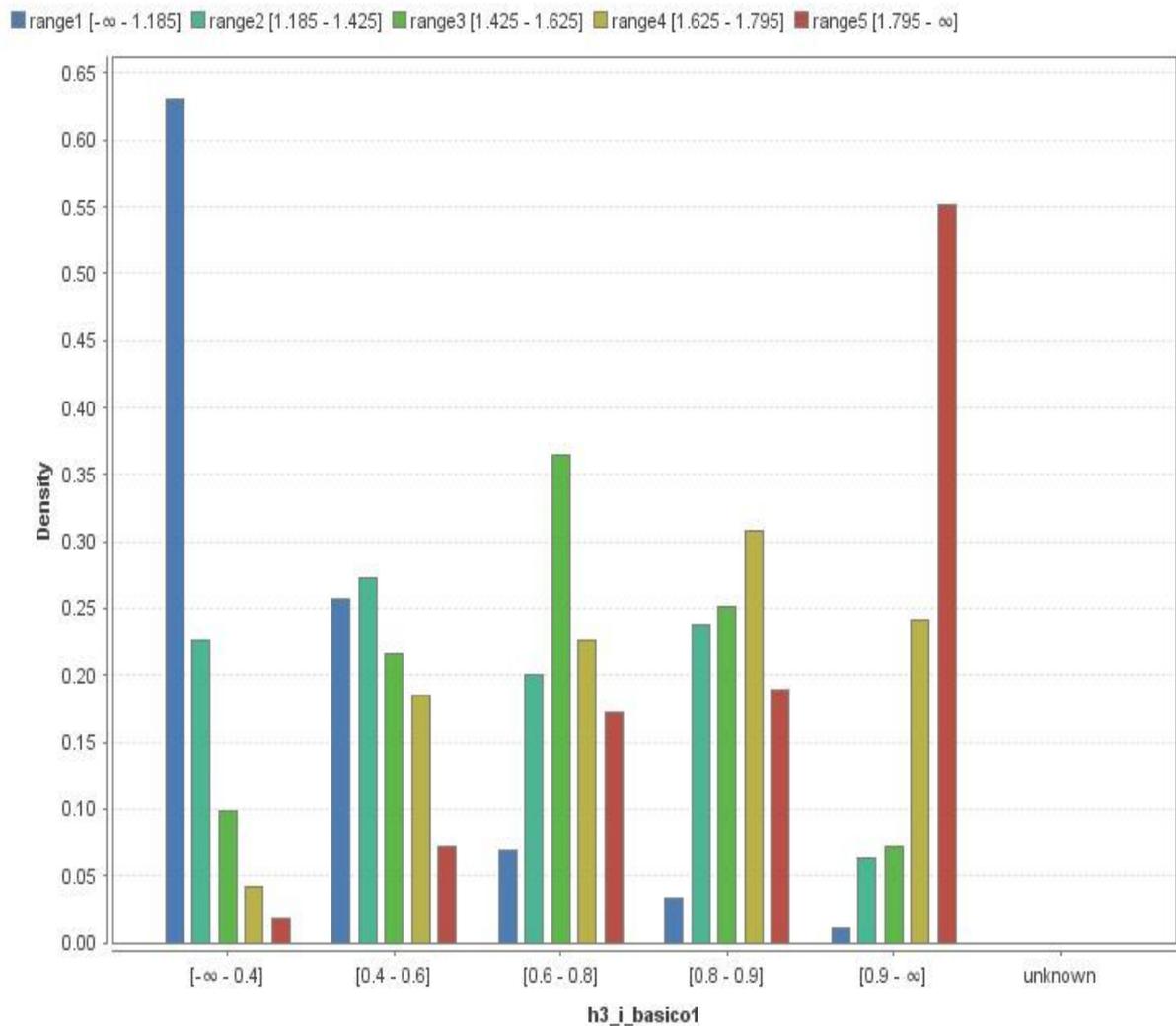
Figura 27 - Naive Bayes fase B.



Fonte: Autor, (2019).

Ao se analisar o gráfico da figura 28, que mostra a densidade de probabilidade de ocorrência dos cinco intervalos discretizados do THDv em relação a corrente do 3º harmônico do Básico 1, pode-se verificar que há uma probabilidade de 63% do THDv apresentar um valor igual ou maior que 1,185% quando a corrente do 3º harmônico do Básico 1 apresentar um valor igual ou maior que 0,4%.

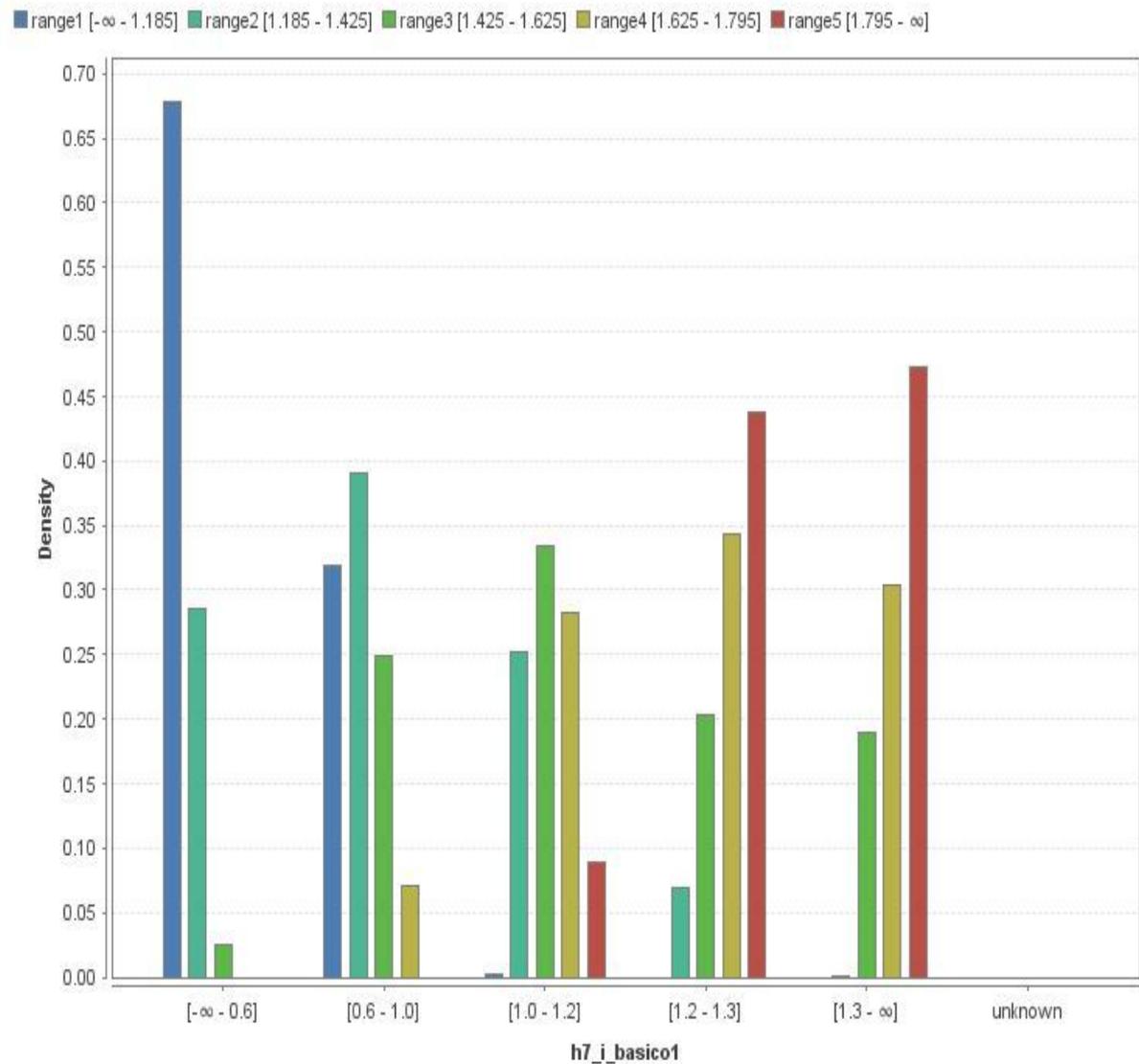
Figura 28 - Densidade de probabilidade dos intervalos discretizados do THDv em relação a corrente do 3º harmônico.



Fonte: Autor, (2019).

Ao se analisar o gráfico da figura 29, que mostra a densidade de probabilidade de ocorrência dos cinco intervalos discretizados do THDv em relação a corrente do 7º harmônico do Básico 1, pode-se verificar que há uma probabilidade de 67,8% do THDv apresentar um valor igual ou maior que 0,6% quando a corrente do 7º harmônico do Básico 1, conforme pode ser visto a seguir na figura 29.

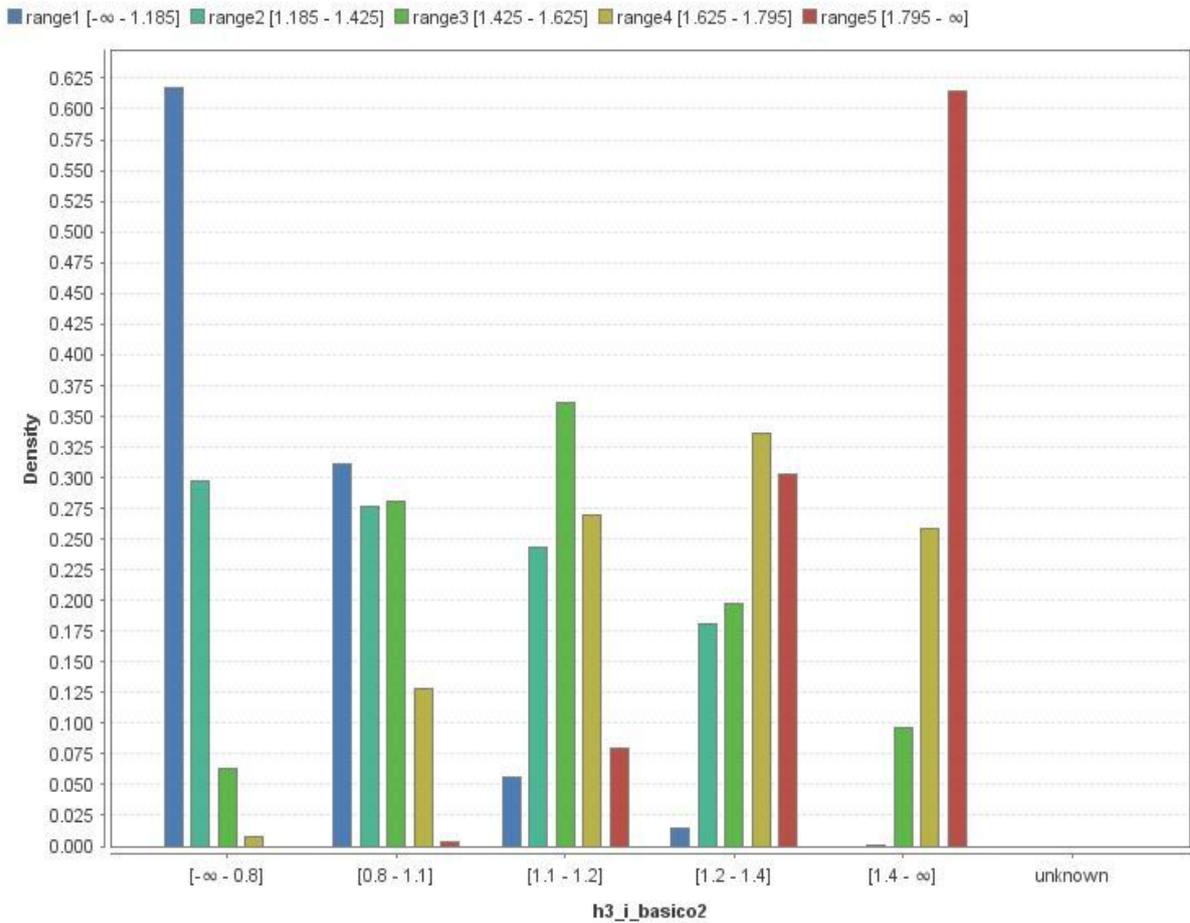
Figura 29 - Densidade de probabilidade dos intervalos discretizados do THDV em relação a corrente do 7º harmônico.



Fonte: Autor, (2019).

Ao se analisar o gráfico da figura 30, que mostra a densidade de probabilidade de ocorrência dos cinco intervalos discretizados do THDV em relação a corrente do 3º harmônico do Básico 2, pode-se verificar que há uma probabilidade de quase 61,7% do THDV apresentar um valor igual ou maior que 1,185% quando a corrente do 3º harmônico do Básico 2 apresentar um valor igual ou maior que 0,8%.

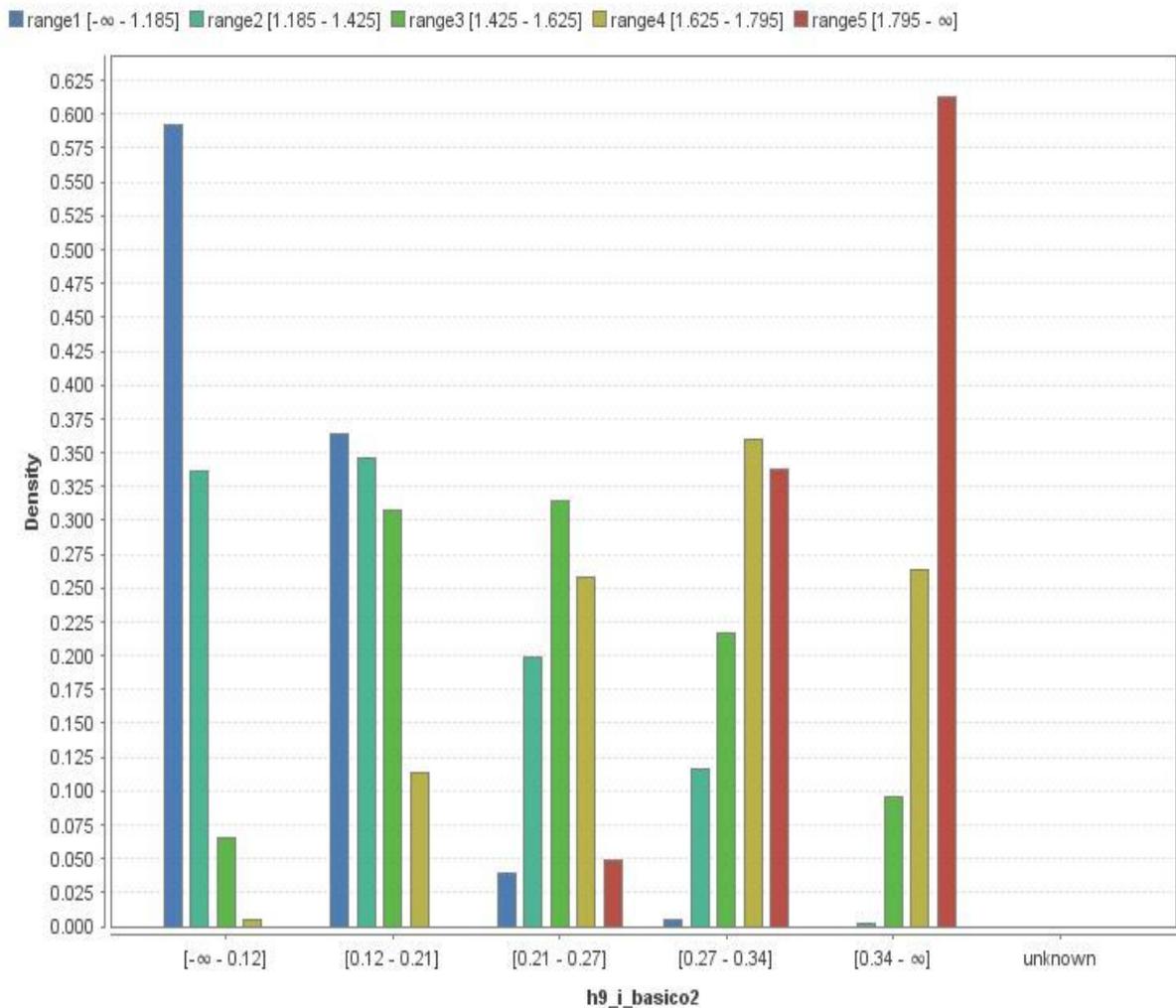
Figura 30 - Densidade de probabilidade dos intervalos discretizados do THDv em relação a corrente do 3º harmônico.



Fonte: Autor, (2019).

Ao se analisar o gráfico da figura 31, que mostra a densidade de probabilidade de ocorrência dos cinco intervalos discretizados do THDv em relação a corrente do 9º harmônico do Básico 2, pode-se verificar que há uma probabilidade de quase 61,3% do THDv apresentar um valor igual ou maior que 1,795% quando a corrente do 9º harmônico do Básico 2 apresentar um valor igual ou maior que 0,34%.

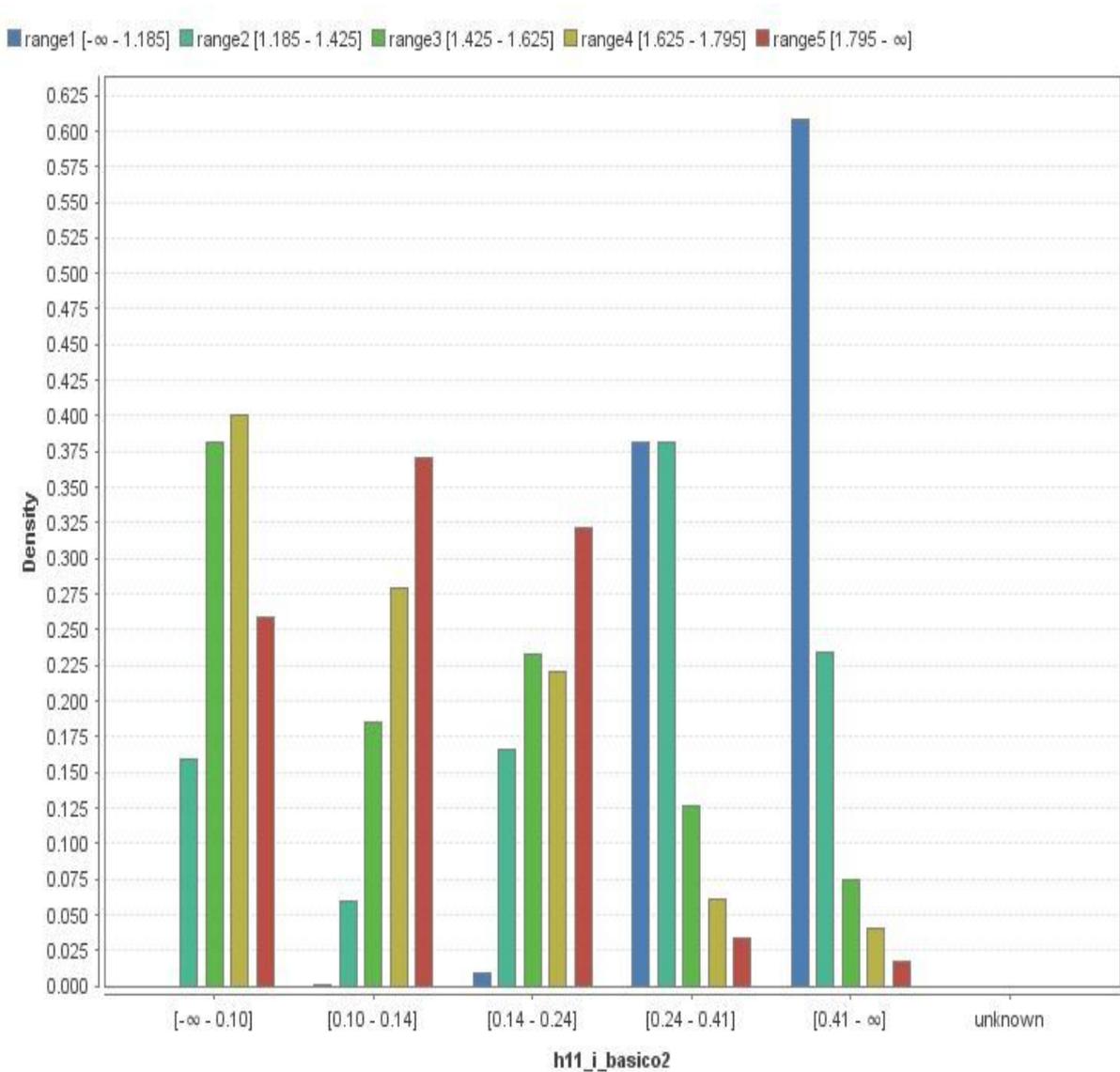
Figura 31 - Densidade de probabilidade dos intervalos discretizados do THDv em relação a corrente do 9º harmônico.



Fonte: Autor, (2019).

Ao se analisar o gráfico da figura 32, que mostra a densidade de probabilidade de ocorrência dos cinco intervalos discretizados do THDv em relação a corrente do 11º harmônico do Básico 2, pode-se verificar que há uma probabilidade de quase 60,8% do THDv apresentar um valor igual ou maior que 1,795% quando a corrente do 11º harmônico do Básico 2 apresentar um valor igual ou maior que 0,41%.

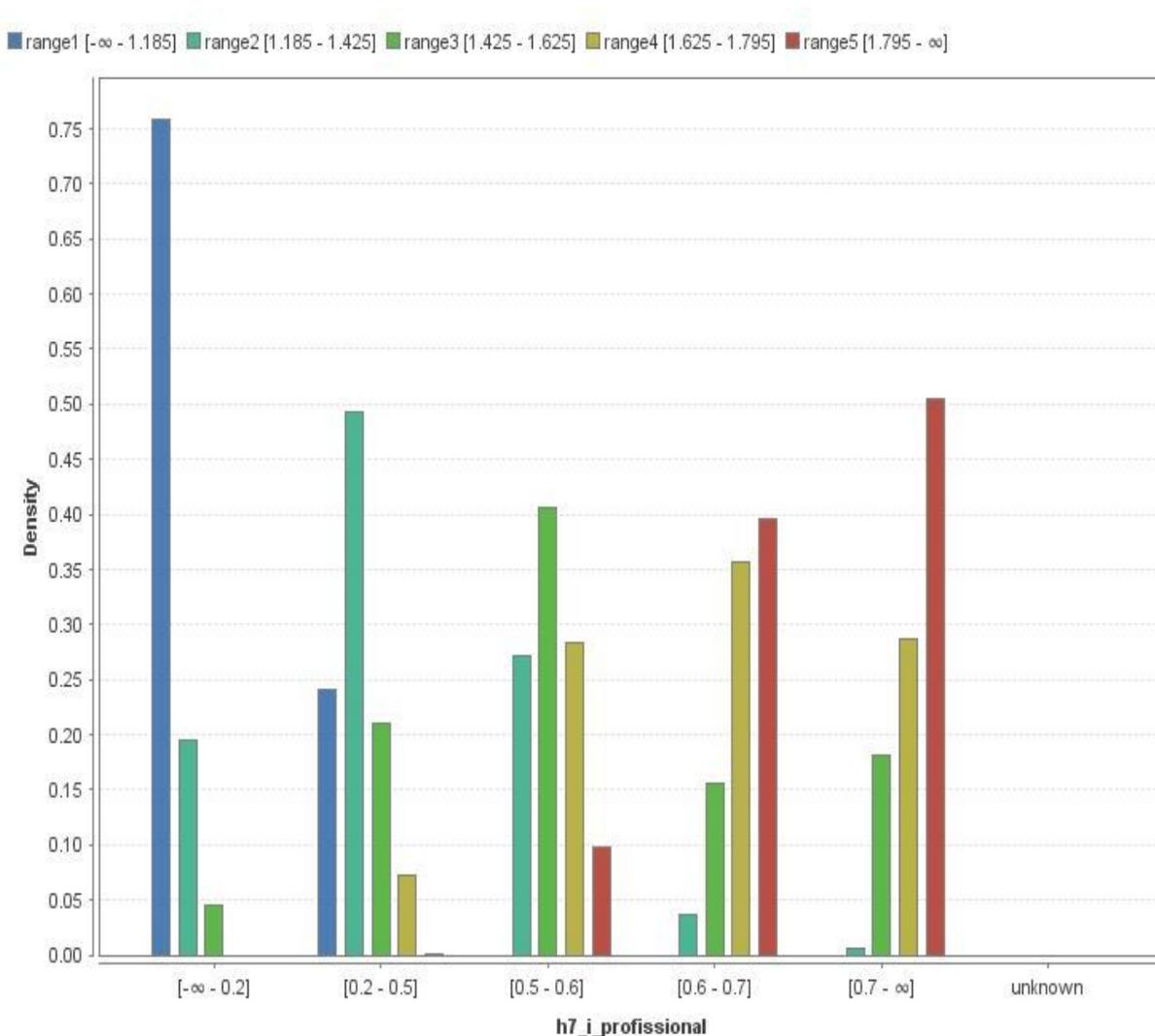
Figura 32 - Densidade de probabilidade dos intervalos discretizados do THDv em relação a corrente do 11º harmônico.



Fonte: Autor, (2019).

Ao se analisar o gráfico da figura 33, que mostra a densidade de probabilidade de ocorrência dos cinco intervalos discretizados do THDv em relação a corrente do 7º harmônico do Profissional, pode-se verificar que há uma probabilidade de quase 60,8% do THDv apresentar um valor igual ou maior que 1,185% quando a corrente do 7º harmônico do Profissional, apresentar um valor igual ou maior que 0,2%.

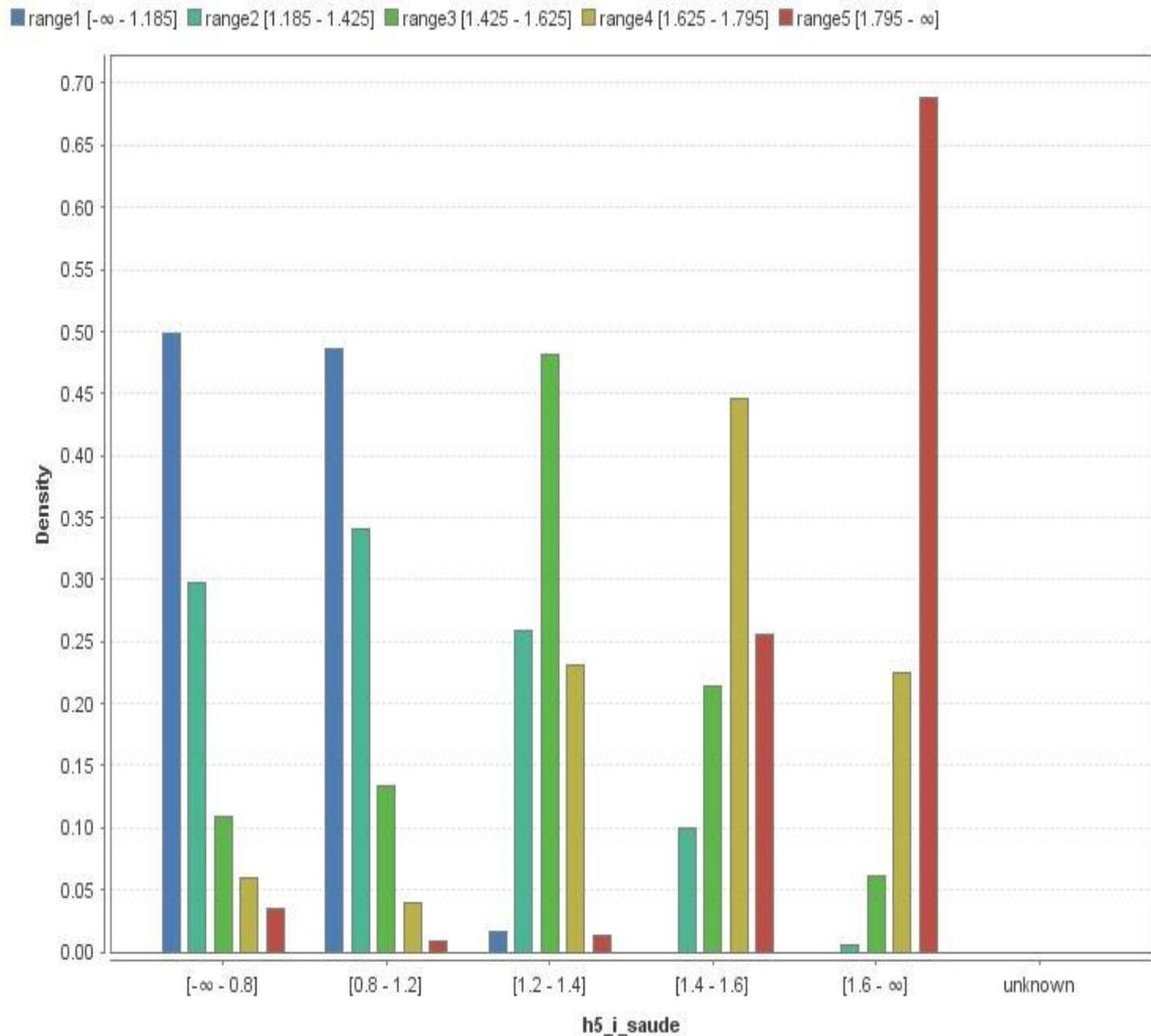
Figura 33 - Densidade de probabilidade dos intervalos discretizados do THDv em relação a corrente do 7º harmônico.



Fonte: Autor, (2019).

Ao se analisar o gráfico da figura 34, que mostra a densidade de probabilidade de ocorrência dos cinco intervalos discretizados do THDv em relação a corrente do 5º harmônico da Saúde, pode-se verificar que há uma probabilidade de quase 68,8% do THDv apresentar um valor igual ou maior que 1,795% quando a corrente do 5º harmônico da Saúde, apresentar um valor igual ou maior que 1,6%.

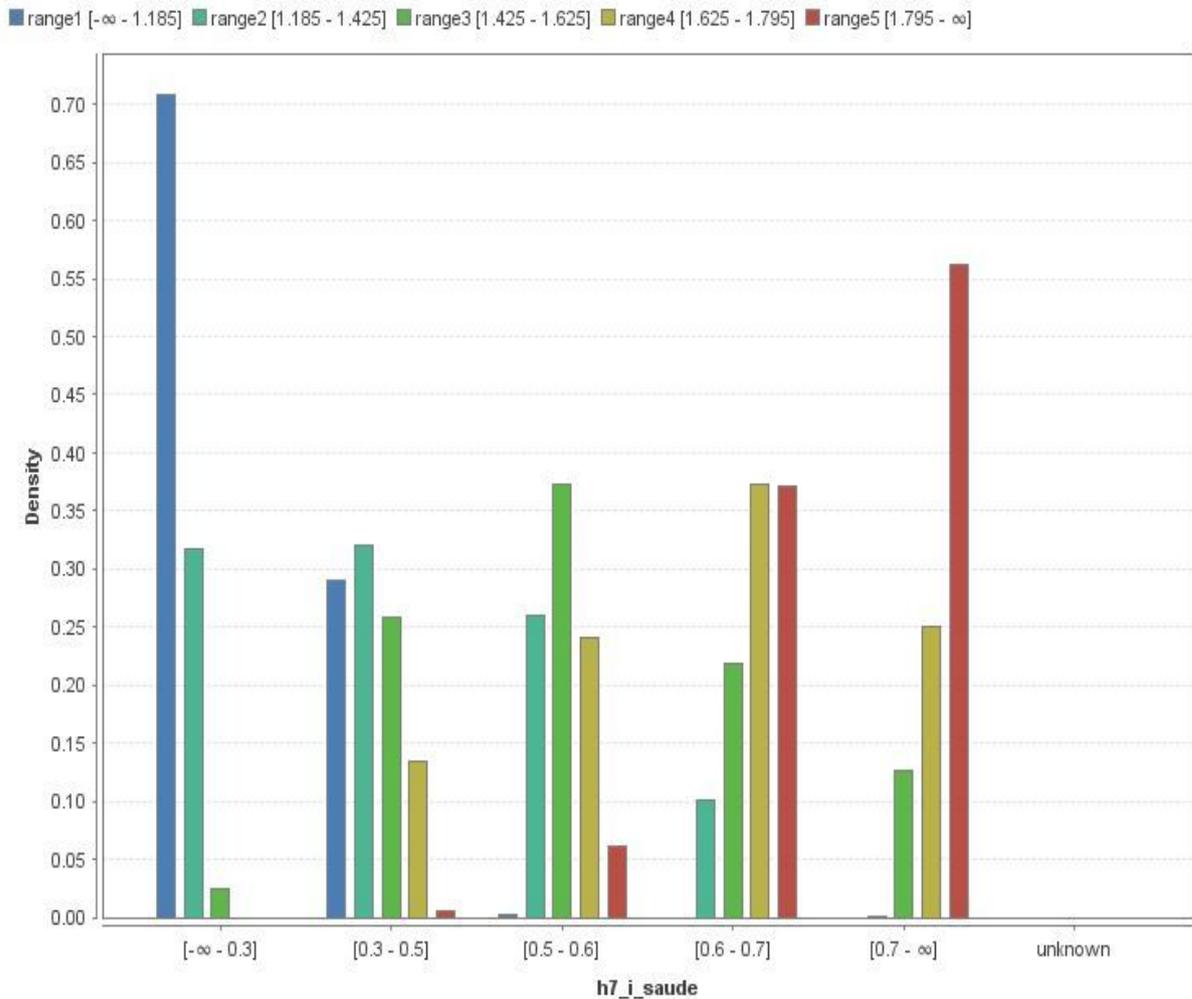
Figura 34 - Densidade de probabilidade dos intervalos discretizados do THDV em relação a corrente do 5º harmônico.



Fonte: Autor, (2019).

Ao se analisar o gráfico da figura 35, que mostra a densidade de probabilidade de ocorrência dos cinco intervalos discretizados do THDV em relação a corrente do 7º harmônico da Saúde, pode-se verificar que há uma probabilidade de quase 70,8% do THDV apresentar um valor igual ou maior que 1,185% quando a corrente do 7º harmônico da Saúde, apresentar um valor igual ou maior que 0,3%.

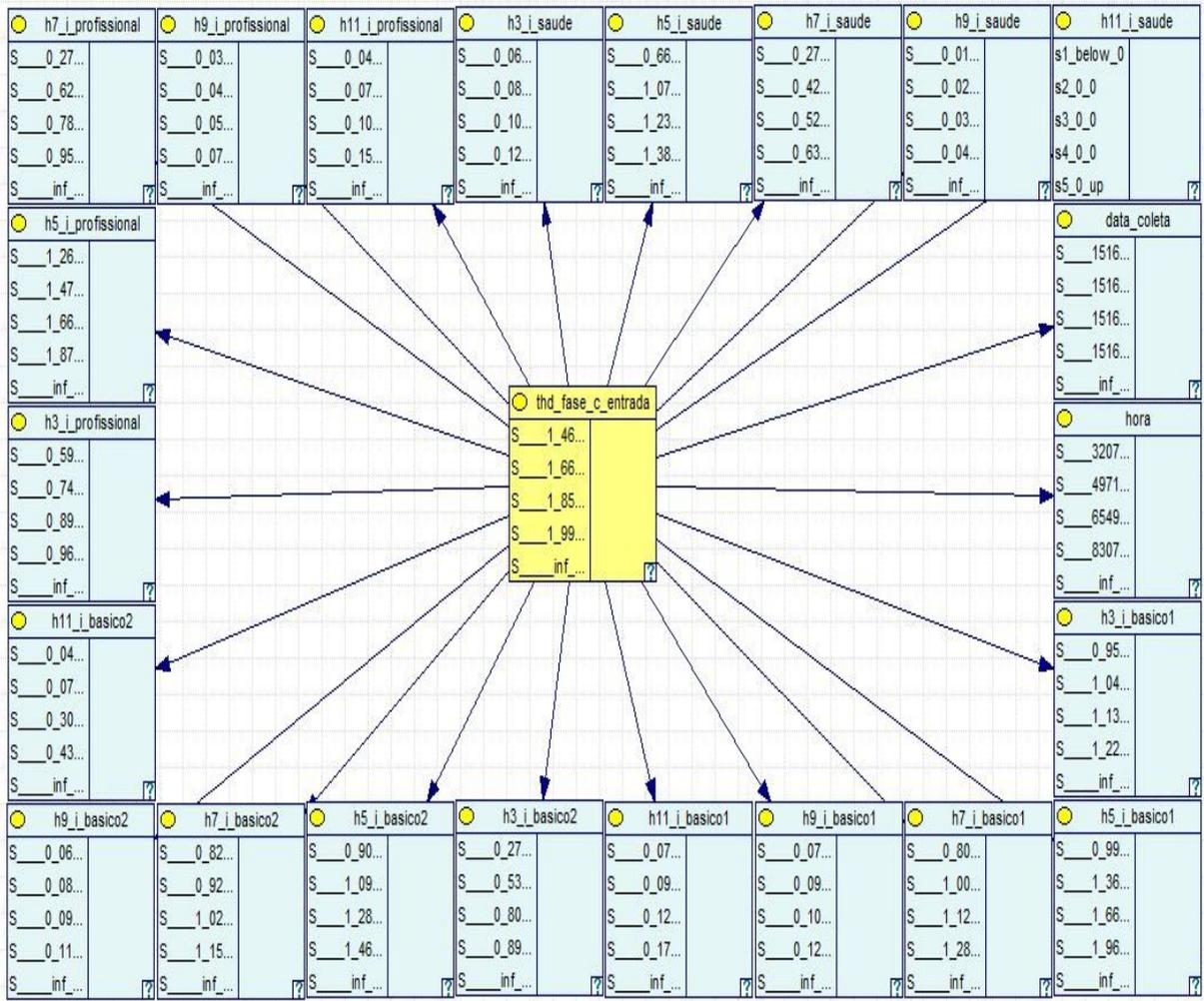
Figura 35 - Densidade de probabilidade dos intervalos discretizados do THDv em relação a corrente do 7º harmônico.



Fonte: Autor, (2019).

A figura 36 apresenta as probabilidades obtidas com a criação da rede Naive Bayes para a fase C, criada com os dados provenientes da campanha de medição realizada nas instalações da UFPA, cujas grandezas registradas foram: taxa de distorção harmônica total de tensão (THDv) da Cabine de medição de entrada no setor Básico, e as correntes harmônicas de ordem 3º, 5º, 7º, 9º e 11º dos Campi: Básico 1, Básico 2, Profissional e Saúde, temos a figura 36 a seguir.

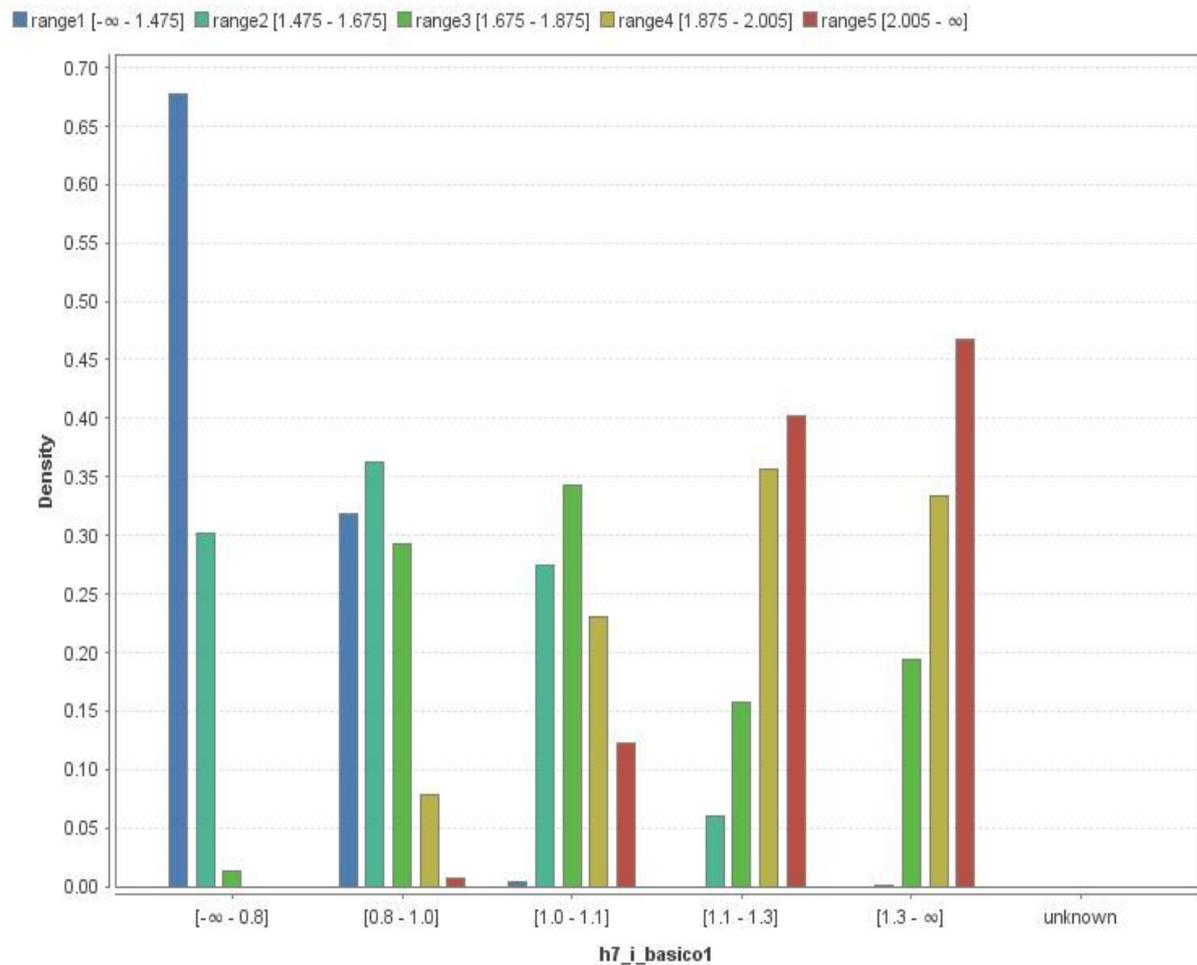
Figura 36 - Naive Bayes fase C.



Fonte: Autor, (2019).

Ao se analisar o gráfico da figura 37, o qual mostra a densidade de probabilidade de ocorrência dos cinco intervalos discretizados do THDv em relação a corrente do 7º harmônico do Básico 1, pode-se verificar que há uma probabilidade de quase 67,7% do THDv apresentar um valor igual ou maior que 1,475% quando a corrente do 7º harmônico do Básico 1, apresentar um valor igual ou maior que 0,8%.

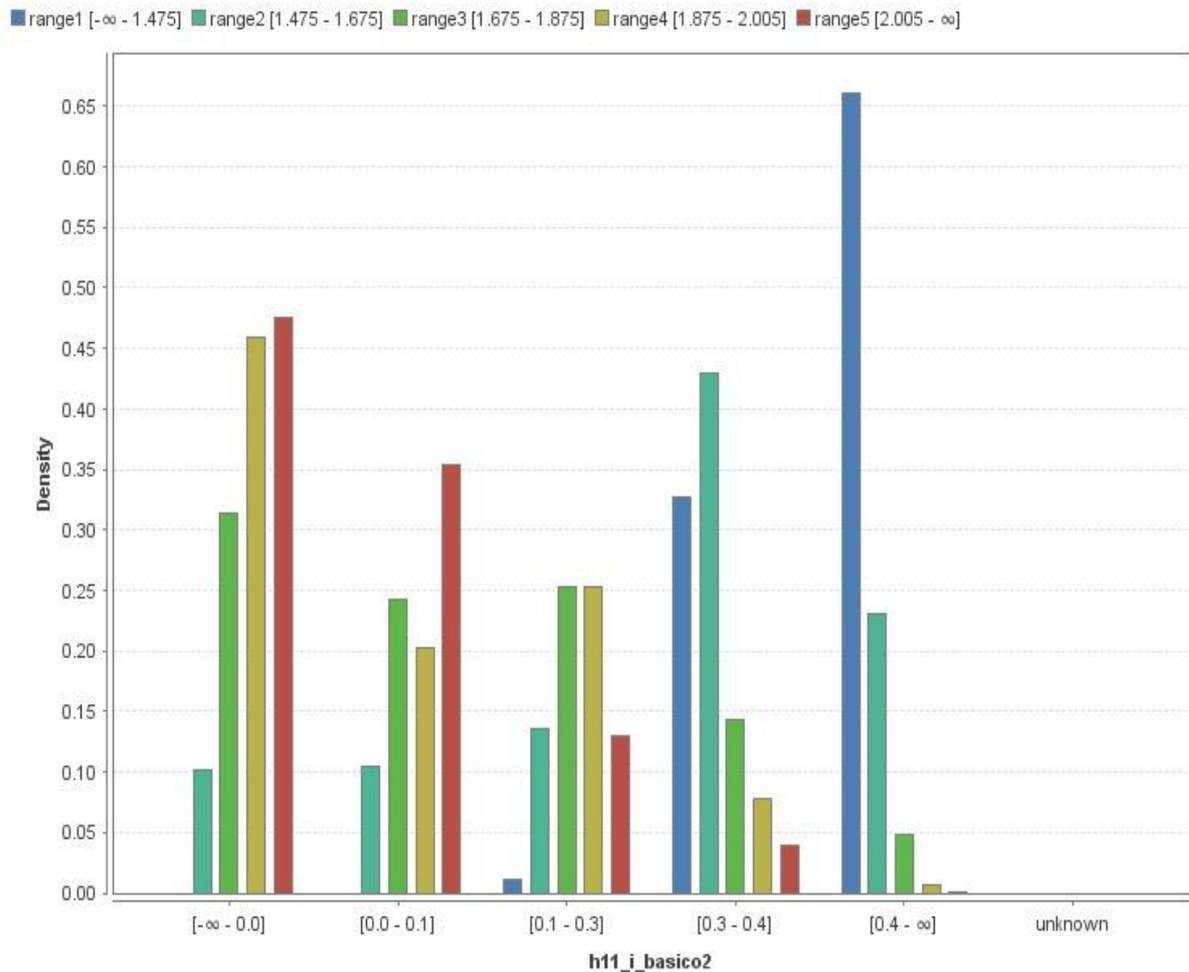
Figura 37 - Densidade de probabilidade dos intervalos discretizados do THDv em relação a corrente do 7º harmônico.



Fonte: Autor, (2019).

Ao se analisar o gráfico da figura 38, que apresenta a densidade de probabilidade de ocorrência dos cinco intervalos discretizados do THDv em relação a corrente do 11º harmônico do Básico 2, pode-se verificar que há uma probabilidade de 83,1% do THDv apresentar um valor igual ou maior que 1,175% quando a corrente do 11º harmônico do Básico 2, apresentar um valor igual ou maior que 0,4 % conforme mostra a figura 38.

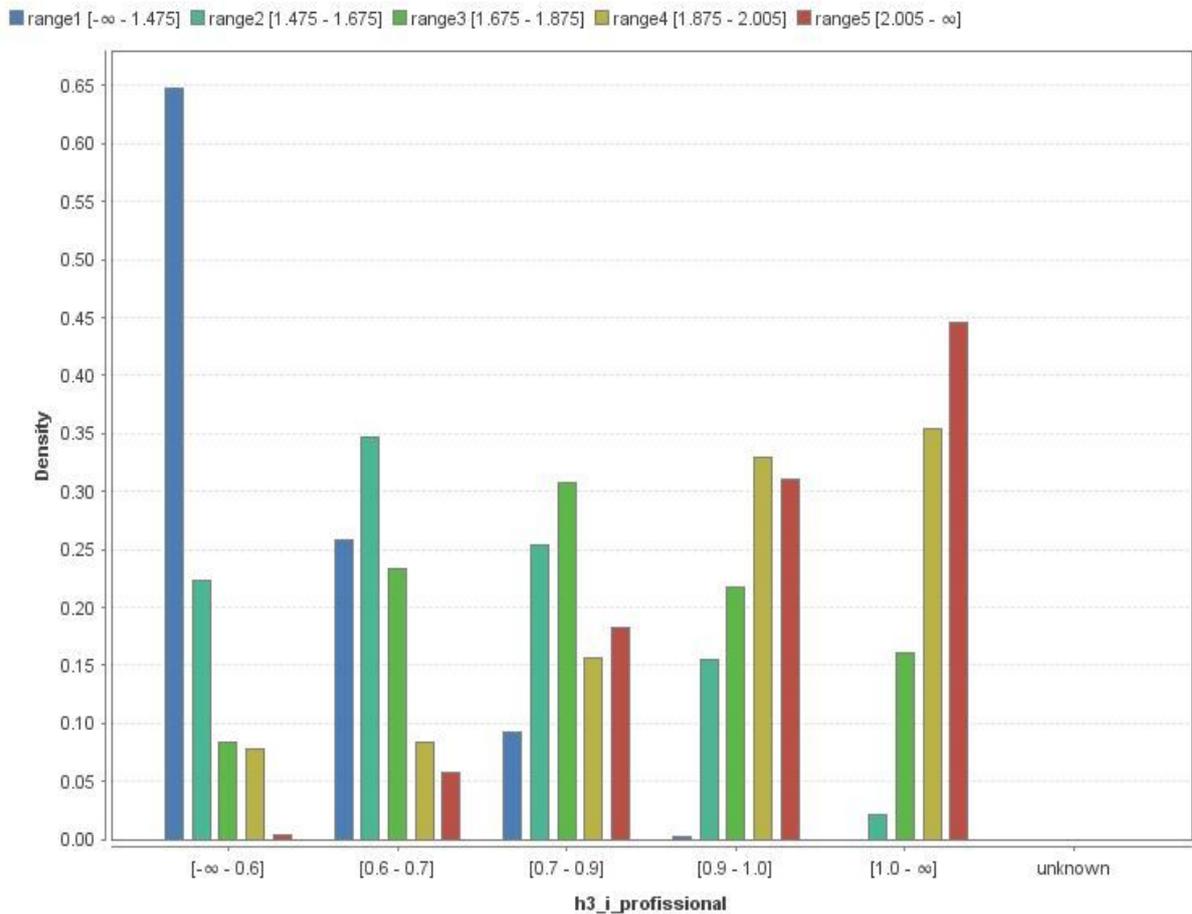
Figura 38 - Densidade de probabilidade dos intervalos discretizados do THDV em relação a corrente do 11º harmônico.



Fonte: Autor, (2019).

Ao se analisar o gráfico da figura 39, que mostra a densidade de probabilidade de ocorrência dos cinco intervalos discretizados do THDV em relação a corrente do 3º harmônico do Profissional, pode-se verificar que há uma probabilidade de 64,7% do THDV apresentar um valor igual ou maior que 1,475% quando a corrente do 3º harmônico do Profissional apresentar um valor igual ou maior que 0,6% conforme mostra a figura 39.

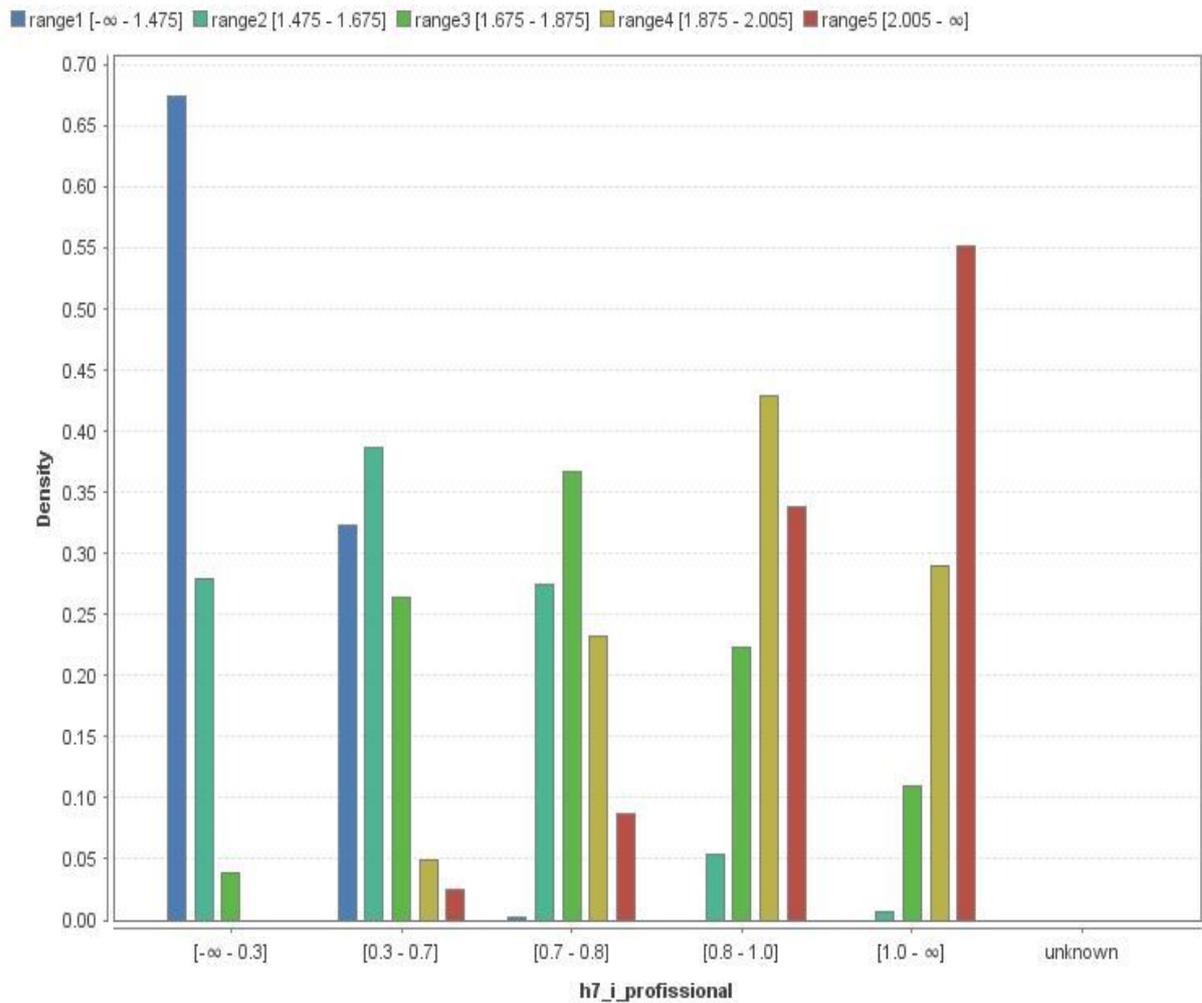
Figura 39 - Densidade de probabilidade dos intervalos discretizados do THDv em relação a corrente do 3º harmônico.



Fonte: Autor, (2019).

Ao se analisar o gráfico da figura 40, que mostra a densidade de probabilidade de ocorrência dos cinco intervalos discretizados do THDv em relação a corrente do 7º harmônico do Profissional, pode-se verificar que há uma probabilidade de 67,4% do THDv apresentar um valor igual ou maior que 1,475% quando a corrente do 7º harmônico do Profissional apresentar um valor igual ou maior que 0,3% conforme mostra a figura 40.

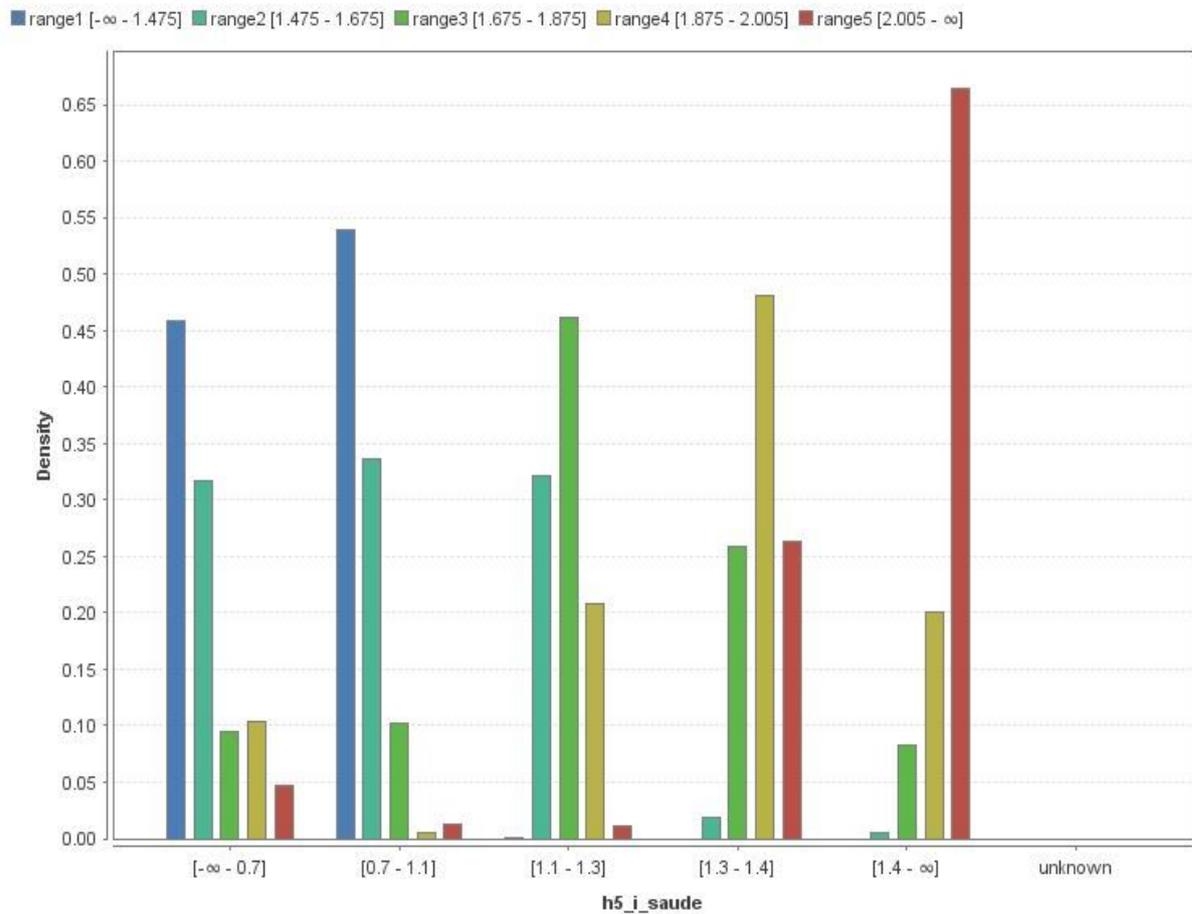
Figura 40 - Densidade de probabilidade dos intervalos discretizados do THDv em relação a corrente do 7º harmônico.



Fonte: Autor, (2019).

Ao se analisar o gráfico da figura 41, que mostra a densidade de probabilidade de ocorrência dos cinco intervalos discretizados do THDv em relação a corrente do 5º harmônico da Saúde, pode-se verificar que há uma probabilidade de 66,4% do THDv apresentar um valor igual ou maior que 2,005% quando a corrente do 5º harmônico da Saúde, apresentar um valor igual ou maior que 1,4% conforme mostra a figura 41.

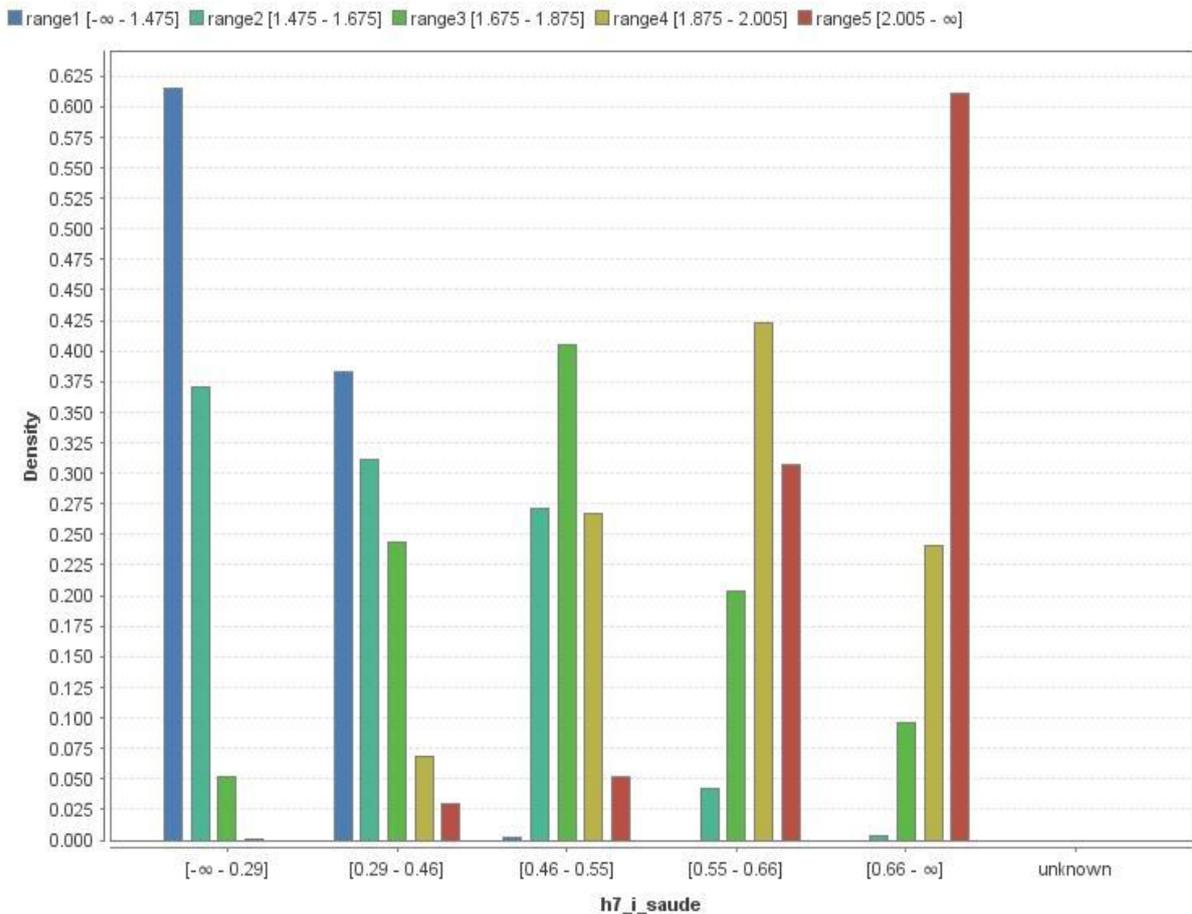
Figura 41 - Densidade de probabilidade dos intervalos discretizados do THDv em relação a corrente do 5º harmônico.



Fonte: Autor, (2019).

Ao se analisar o gráfico da figura 42, que mostra a densidade de probabilidade de ocorrência dos cinco intervalos discretizados do THDv em relação a corrente do 7º harmônico da Saúde, pode-se verificar que há uma probabilidade de 61,5% do THDv apresentar um valor igual ou maior que 1,1475% quando a corrente do 7º harmônico da Saúde, apresentar um valor igual ou maior que 0,29% conforme mostra a figura 42.

Figura 42 - Densidade de probabilidade dos intervalos discretizados do THD_v em relação a corrente do 7º harmônico.



Fonte: Autor, (2019).

6.4 Considerações Finais

Abordou-se neste capítulo as análises realizadas nos dados com a aplicação do processo de KDD, onde foram realizadas as fases de compreensão do domínio, seleção de dados, pré-processamento, transformação, objetivos do processo KDD, seleção de modelo (Naive Bayes), mineração de dados, interpretação e avaliação dos resultados do processo, que ocorreu com a aplicação da técnica de Naive Bayes como técnica de aprendizado, atuando no conhecimento descoberto com objetivo de extrair conhecimento das bases de dados dos processos monitorados para que auxiliem na tomada de decisão com relação ao conteúdo harmônico presentes nos nas áreas monitoradas da UFPA campus Guamá.

CAPÍTULO VII

CONSIDERAÇÕES FINAIS

7.1 CONCLUSÕES

O presente estudo apresenta um conjunto de etapas que permitem o monitoramento e a avaliação da qualidade de energia através da análise do indicador taxa de distorção harmônica (THD), foi realizado no sistema elétrico da Universidade Federal do Pará, Campus Universitário do Guamá que tem o seu fornecimento de energia elétrica em 13.8 kV, através do alimentador Guamá GM-06, proveniente da Subestação Guamá da concessionária de energia local. Todos os seus prédios são atendidos pelos circuitos oriundos da cabine de medição geral, localizada no Campus Básico, onde é instalada a proteção primária geral. A utilização de técnicas de inteligência computacional, e a utilização dos processos de KDD que agrega a visão do especialista no domínio, baseado na recomendação do PRODIST modulo 8, revisão 10, contribui para uma avaliação do diagnóstico de forma eficiente contemplando as recomendações regulamentadoras vigentes.

Esta pesquisa agrega um conhecimento importante sobre qualidade de energia elétrica, ou seja, os indicadores utilizados, tendo como base as normas e recomendações nacionais e internacionais tornam o estudo uma fonte de consulta, para os profissionais da área, muito importante.

Contempla as características, conceitos, fases e algoritmo utilizado no processo KDD, bem como as etapas de monitoramento e a avaliação da QEE fundamentos de acordo com PRODIST no módulo 8 Revisão 10, Normas IEC e IEEE, através dos indicadores de taxa de distorção harmônica total (THD) e as distorções harmônicas individuais de 3°, 5°, 7°, 9° e 11° ordens; Demonstrando as aplicações da inteligência computacional nas análises de fenômenos de diversas áreas do conhecimento, dando maior ênfase na análise realizada na Universidade Federal do Pará Campus Guamá, onde foi montado um estudo de QEE e inteligência computacional com objetivo de permitir uma tomada de decisão no cenário atual, mediante os dados extraídos dos campis da referida instituição.

Portanto, ao avaliar o conteúdo de impactos harmônicos existente na rede elétrica de um Campus Universitário, por meio de parâmetros obtidos com medições feitas por equipamentos devidamente calibrados, utilizando técnicas estatísticas e mineração de dados

com base nas análises realizadas utilizando o algoritmo de *Naive Bayes* em uma base discretizada por frequência obteve-se as seguintes conclusões: A curva ROC apresentou os seguintes resultados para as fases A igual 0,998, fase B 0,982 e fase B 0,981 e no contexto geral as harmônicas contribuíram entre 60 e 75% no impacto harmônico:

Quando o intervalo discretizado do THD variou de $[-\infty - 1.2]$ os harmônicos apresentaram os seguintes valores mais significativos para os intervalos discretizados de corrente, 0,70% h9_basico2 e 0,67% no h7_saúde da fase A.

Quando o intervalo discretizado do THD variou de $[-\infty - 1.185]$ os harmônicos apresentaram os seguintes valores mais significativos para os intervalos discretizados de corrente, 0,75% h7_profissional e 0,70% no h7_saúde da fase B.

Quando o intervalo discretizado do THD variou de $[-\infty - 1.475]$ os harmônicos apresentaram os seguintes valores mais significativos para os intervalos discretizados de corrente, 0,67% h7_básico1 e 0,77% no h7_saúde da fase C.

O que reforça a preocupação das recomendações das normas internacionais e do PRODIST no módulo 8, revisão 10, o qual referência a qualidade de energia em sistemas elétricos.

Dentre as principais contribuições deste trabalho:

- Apresentar um estudo sobre a qualidade de energia elétrica, tendo como base o indicador, a taxa de distorção harmônica (THD) usando as técnicas de KDD e classificador Naive Bayes;
- Analisar as áreas envolvidas nos processos de uma universidade, viabilizar uma avaliação do conteúdo harmônico presente em cada um dos campi, analisar os maiores impactos de harmônicos individuais para cada local;
- Fornecer informações com base nos dados analisados que auxiliem na tomada de decisão, controle dos harmônicos gerados em cada área da universidade, evitando penalidades devido a geração de harmônicos fora dos limites estabelecidos nas normas vigentes das agências reguladoras.

7.2 SUGESTÕES PARA TRABALHOS FUTUROS

Conforme explicados durante a dissertação podem ser consideradas as seguintes recomendações para trabalhos posteriores, haja vista que o tema é relevante para área de

engenharia elétrica e de computação como ferramentas importantes para o desenvolvimento das análises da qualidade de energia:

- Utilizar outras técnicas de Inteligência Computacional para a avaliação da qualidade de energia, tais como: Arvore de decisão, Redes Neurais, Algoritmos Genéticos etc;
- Compara o desempenho de duas ou mais técnicas de mineração de dados na avaliação da qualidade de energia;
- Nesta área, ainda há muito a ser desenvolvido e consolidado, sugerimos a inclusão de outras grandezas elétricas;
- Continuar trabalhando em modelos de cargas não lineares para determinar mais precisamente os harmônicos gerados e os possíveis efeitos na rede elétrica.

REFERÊNCIAS

AL-BATAINEH, H.; KAVASSERI, R. G. Islanding Detection with Data Mining Methods-A Comparative Study. Green Technologies Conference (GreenTech), 2017 Ninth Annual IEEE, 2017. IEEE. p.104-109.

AL-GARADI, M. A. et al. Using online social networks to track a pandemic: A systematic review. **Journal of Biomedical Informatics**, v. 62, p. 1-11, 8// 2016. ISSN 1532-0464. Disponível em: < <http://www.sciencedirect.com/science/article/pii/S1532046416300351> >.

ANEEL-PRODIST. Procedimentos de Distribuição de Energia Elétrica no Sistema Elétrico Nacional – PRODIST Módulo 8 – Qualidade da Energia Elétrica. **Agência Nacional de Energia Elétrica**, 01/01/2017 2017.

APTÉ, C. Data mining: an industrial research perspective. **IEEE Computational Science and Engineering**, v. 4, n. 2, p. 6-9, 1997. ISSN 1070-9924.

ASHEIBI, A.; STIRLING, D.; ROBINSON, D. Identification of Load Power Quality Characteristics using Data Mining. 2006 Canadian Conference on Electrical and Computer Engineering, 2006. May 2006. p.157-162.

AZIM, R. et al. A comparative analysis of intelligent classifiers for passive islanding detection in microgrids. PowerTech, 2015 IEEE Eindhoven, 2015. IEEE. p.1-6.

BRAMER, M. **Principles of data mining**. Springer, 2016. ISBN 1447173074.

CASTILLO, G. Métodos de Avaliação para Classificação Supervisada. 2012. Disponível em: < <https://sites.google.com/site/gladyscjaprendizagem/program/bayesian-network-classifiers> >. Acesso em: 05/07/2017.

CESAR, D. G.; VALDOMIRO, V. G.; GABRIEL, O. P. Automatic Power Quality Disturbances Detection and Classification Based on Discrete Wavelet Transform and Artificial Intelligence. 2006 IEEE/PES Transmission & Distribution Conference and Exposition: Latin America, 2006. 15-18 Aug. 2006. p.1-6.

CHENINE, M.; NORDSTROM, L.; JOHNSON, P. Factors in Assessing Performance of Wide Area Communication Networks for Distributed Control of Power Systems. 2007 IEEE Lausanne Power Tech, 2007. 1-5 July 2007. p.1682-1687.

D'ANTONA, G.; MUSCAS, C.; SULIS, S. Harmonic source estimation: A new approach for the localization of nonlinear loads. 2008 13th International Conference on Harmonics and Quality of Power, 2008. Sept. 28 2008-Oct. 1 2008. p.1-6.

_____. State Estimation for the Localization of Harmonic Sources in Electric Distribution Systems. **IEEE Transactions on Instrumentation and Measurement**, v. 58, n. 5, p. 1462-1470, 2009. ISSN 0018-9456.

DASH, R.; PARAMGURU, R. L.; DASH, R. Comparative analysis of supervised and unsupervised discretization techniques. **International Journal of Advances in Science and Technology**, v. 2, n. 3, p. 29-37, 2011.

DIAS, G. A. D. **Harmonic as em sistemas industriais**. EDIPUCRS, 1998. Disponível em: < <https://books.google.com.br/books?id=exjirQEACAAJ> >.

DOORSAMY, W.; CRONJE, W. A. Optimisation of shaft voltage based condition monitoring in generators using a Bayesian approach. 7th IET International Conference on Power Electronics, Machines and Drives (PEMD 2014), 2014. 8-10 April 2014. p.1-5.

DUDA, R. O.; HART, P. E.; STORK, D. G. **Pattern classification**. Wiley, New York, 1973.

FAYYAD, U. M. Data mining and knowledge discovery: Making sense out of data. **IEEE Expert: Intelligent Systems and Their Applications**, v. 11, n. 5, p. 20-25, 1996. ISSN 0885-9000.

FERREIRA, V. C. et al. Fault detection and diagnosis for solar-powered Wireless Mesh Networks using machine learning. Integrated Network and Service Management (IM), 2017 IFIP/IEEE Symposium on, 2017. IEEE. p.456-462.

FRAWLEY, W. J.; PIATETSKY-SHAPIRO, G.; MATHEUS, C. An Overview: Knowledge Discovery in Databases. **Knowledge Discovery in Databases**, p. 1-27, 1991.

HASAN, A. N.; EBOULE, P. P.; TWALA, B. The use of machine learning techniques to classify power transmission line fault types and locations. Optimization of Electrical and Electronic Equipment (OPTIM) & 2017 Intl Aegean Conference on Electrical Machines and Power Electronics (ACEMP), 2017 International Conference on, 2017. IEEE. p.221-226.

HUANG, Z. et al. Transient power quality assessment based on big data analysis. 2014 China International Conference on Electricity Distribution (CICED), 2014. 23-26 Sept. 2014. p.1308-1312.

IYER, S. P. et al. Inferring functional connectivity in MRI using Bayesian network structure learning with a modified PC algorithm. **NeuroImage**, v. 75, p. 165-175, 7/15/ 2013. ISSN 1053-8119. Disponível em: < <http://www.sciencedirect.com/science/article/pii/S1053811913001997> >.

JAIWEI, H.; KAMBER, M. Data mining: concepts and techniques. ed: **Morgan Kaufmann San Francisco**, 2006.

KHAN, M. R. et al. Non stationary signal analysis and classification using FFT transform and Naive Bayes classifier. 2015 IEEE Power, Communication and Information Technology Conference (PCITC), 2015. 15-17 Oct. 2015. p.967-972.

KJÆRULFF, U. B.; MADSEN, A. L. **Bayesian Networks and Influence Diagrams: A Guide to Construction and Analysis**. Springer New York, 2013. ISBN 9781461451044. Disponível em: < https://books.google.com.br/books?id=PQN9F5unh_YC >.

KJAERULFF, U. B.; MADSEN, A. L. **Bayesian networks and influence diagrams: a guide to construction and analysis**. Springer Science & Business Media, 2012. ISBN 1461451043.

KOTSIANTIS, S.; KANELLOPOULOS, D. Discretization techniques: A recent survey. **GESTS International Transactions on Computer Science and Engineering**, v. 32, n. 1, p. 47-58, 2006.

LAKSHMI, B.; RAGHUNANDHAN, G. A conceptual overview of data mining. Innovations in Emerging Technology (NCOIET), 2011 National Conference on, 2011. IEEE. p.27-32.

LARRAÑAGA, P.; INZA, I.; MOUJAHID, A. Tema 6. Clasificadores Bayesianos. **Departamento de Ciencias de la Computación e Inteligencia Artificial–Universidad del País Vasco-Euskal Herriko Unibertsitatea**, 1997.

LEÃO, R.; ANTUNES, F.; SAMPAIO, R. **Harmônicos em Sistemas Elétricos**. Elsevier Brasil, 2015. ISBN 9788535274400. Disponível em: < <https://books.google.com.br/books?id=NtNeAwAAQBAJ> >.

LINOFF, G. S.; BERRY, M. J. **Data mining techniques: for marketing, sales, and customer relationship management**. John Wiley & Sons, 2011. ISBN 1118087453.

MAIMON, O.; ROKACH, L. **Data mining and knowledge discovery handbook**. Springer, 2005.

MARKOV, Z.; LAROSE, D. T. **Data mining the Web: uncovering patterns in Web content, structure, and usage**. John Wiley & Sons, 2007. ISBN 0470108088.

MINSKY, M. Steps toward artificial intelligence. **Proceedings of the IRE**, v. 49, n. 1, p. 8-30, 1961. ISSN 0096-8390.

MITCHELL, T. M. **Machine Learning**. McGraw-Hill, 1997. ISBN 9780071154673. Disponível em: < <https://books.google.com.br/books?id=EoYBngEACAAJ> >.

MOCANU, E.; NGUYEN, P. H.; GIBESCU, M. Energy disaggregation for real-time building flexibility detection. Power and Energy Society General Meeting (PESGM), 2016, 2016. IEEE. p.1-5.

MORENO, H. Harmônicas nas Instalações Elétricas: Causas. **Efeitos e Soluções, 1st edn, Procobre-Instituto Brasileiro do Cobre, Sao Paulo, SP, Brazil**, 2001.

MATOS, E. O. DE. **Análise não Paramétrica para Identificação de Fontes de Distorções Harmônicas em Sistemas de Energia Elétrica: Um Estudo Aplicado no Campus Universitário do Guamá da Universidade Federal do Pará**. [s.l.] Universidade Federal do Pará, 2016.

MURTY, M. N.; DEVI, V. S. **Pattern recognition: An algorithmic approach**. Springer Science & Business Media, 2011. ISBN 0857294954.

OLIVEIRA, E. F. D. et al. Voltage THD Analysis Using Knowledge Discovery in Databases with a Decision Tree Classifier. **IEEE Access**, v. PP, n. 99, p. 1-1, 2017.

PYLARINOS, D. et al. Investigation and classification of field leakage current waveforms. **IEEE Transactions on Dielectrics and Electrical Insulation**, v. 19, n. 6, p. 2111-2118, 2012. ISSN 1070-9878.

R., K.; F., P. Glossary of Terms Special Issue on Applications of Machine Learning and the Knowledge Discovery Process. Disponível em: < <http://robotics.stanford.edu/~ronnyk/glossary.html> >. Acesso em: 03/07/2017 18:58.

RAMOS, G.; TORRES, A.; ROGNON, J.-P. Análisis de confiabilidad de sistemas industriales aplicando redes bayesianas considerando aspectos de PQ y seguridad. **Ieee latin america transactions**, v. 5, n. 8, 2007.

SAGIROGLU, S. et al. Development of an Intelligent Decision Support System for Determining the Efficiency of Shunt Active Power Filter. 2009 International Conference on Machine Learning and Applications, 2009. 13-15 Dec. 2009. p.626-631.

SANTANA, Á. L. D. **Projeto e implementação de um sistema de suporte à decisão para o observatório de saúde da Amazônia**. 2005. Dissertação].[Pará]: Universidade Federal do Pará

SOMASUNDARAM, G. et al. Classification and Counting of Composite Objects in Traffic Scenes Using Global and Local Image Analysis. **IEEE Transactions on Intelligent Transportation Systems**, v. 14, n. 1, p. 69-81, 2013. ISSN 1524-9050.

SILVA et al., *Naive Bayes Aplicados na Análise de Impactos Harmônicos em Sistemas Elétricos Industriais*, XI IEEE/IAS International Conference on Industry Applications – INDUSCON, Brasil, 2014.

SUMATHI, S.; SIVANANDAM, S. **Introduction to data mining and its applications**. Springer, 2006. ISBN 3540343512.

SUPRIYA, P.; NAMBIAR, T. N. P. Harmonic State Estimation for a simple power system model using Independent Component Analysis. 2015 IEEE 9th International Conference on Intelligent Systems and Control (ISCO), 2015. 9-10 Jan. 2015. p.1-4.

TAMANO, K.; TSUJI, H. Relation of home energy consumption and static properties of consumers. 2011 3rd International Conference on Awareness Science and Technology (iCAST), 2011. 27-30 Sept. 2011. p.215-220.

TAN, P. N. et al. **Introdução ao datamining: mineração de dados**. Ciencia Moderna, 2009. ISBN 9788573937619. Disponível em: < <https://books.google.com.br/books?id=69d6PgAACAAJ> >.

TORRES, A.; RUEDA, M. T.; REYES, D. Bayesian Networks for Power Quality Analysis in the Industrial Sector. 2006 International Conference on Probabilistic Methods Applied to Power Systems, 2006. 11-15 June 2006. p.1-7.

UZUNOGLU, B. **Locating distribution power system fault employing Bayes theorem with subjective logic**. 2017.

ZHANG, H.; LU, Y. Learning Bayesian network classifiers from data with missing values. TENCON '02. Proceedings. 2002 IEEE Region 10 Conference on Computers, Communications, Control and Power Engineering, 2002. 28-31 Oct. 2002. p.35-38 vol.1.